Agency for
Science, Technology
and Research
SINGAPORE

Bioinformatics Institute (BII)

Protein Sequence Analysis

Executive Director
MAURER-STROH Sebastian

Research Scientist
LIMVIPHUVADH Vachiranee

Senior Post-Doctoral Research Fellow
KENANOV Dimitar

Senior Post-Doctoral Research Fellow
HO Wei-Hao Joses

Senior Post-Doctoral Research Fellow
MAK Tze Minn Sandy

Post-Doctoral Research Fellow
CHONG Cheng Shoong Ken

Research Manager
LEE Tze Chuan Raphael

Senior Research Officer
XU Yani Angela

Research Officer
MIYAJIMA Jhoann

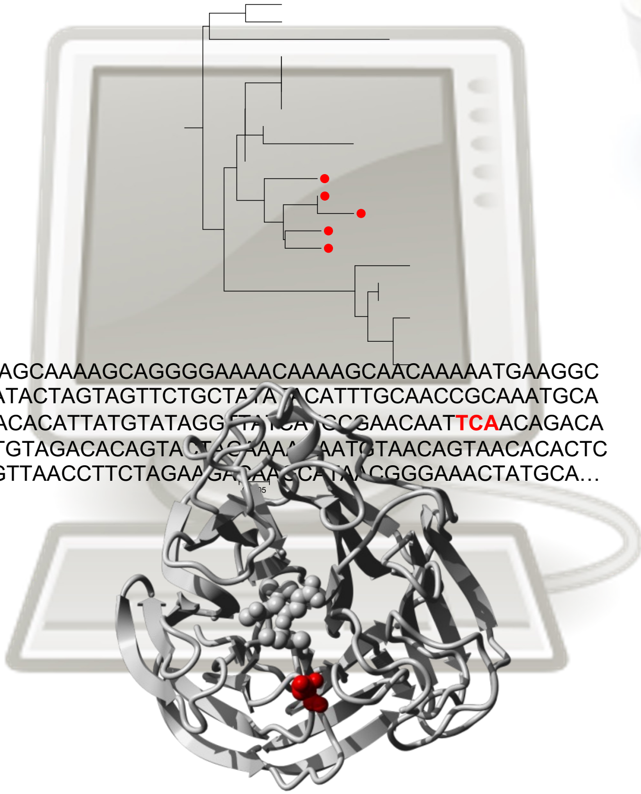
Research Officer
CHEW Yi Hong

Research Officer
MAKHEJA Meera

CREATING GROWTH, ENHANCING LIVES

Viruses

Computational Sequence and Structure Analysis



...AGCAAAAGCAGGGGAAAACAAAAGCAACAAAAATGAAGGC
AATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCA
GACACATTATGTATAGGTTATCATGCGAACAATTCAACAGACA
CTGTAGACACAGTACTAGAAAAGAATGTAACAGTAACACACTC
TGTTAACCTTCTAGAAGACAAGCATAACGGGAAACTATGCA...

Food flavors

Drugs

Human variation

We are working at the interface between sequence and structure

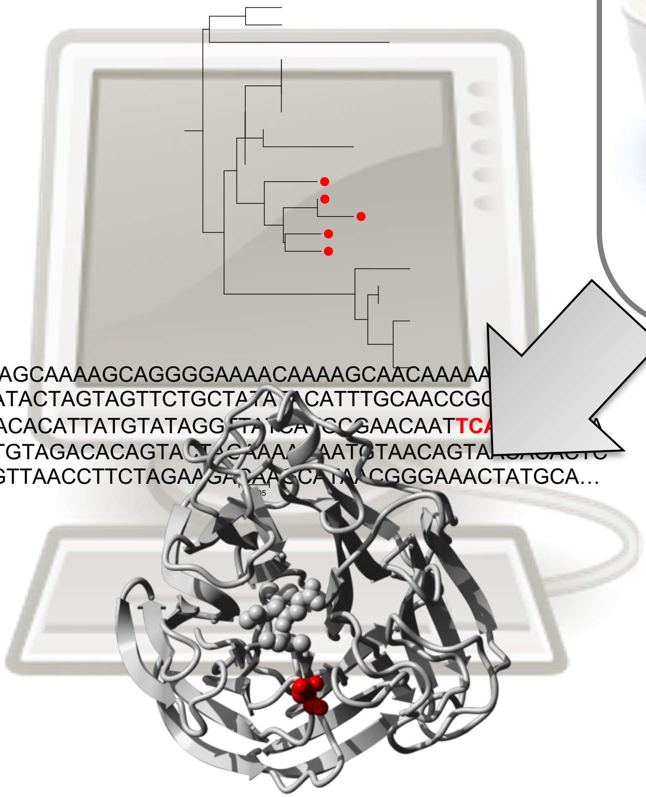
Product Safety

head&shoulders
green apple
ADVANCED FORMULA
head&shoulders
conditioner
green apple

CREATING GROWTH, ENHANCING LIVES

Viruses

Computational Sequence and Structure Analysis


...AGCAAAAGCAGGGGAAAACAAAAGCAACAAAAA
AATACTAGTAGTTCTGCTATA ACATTTGCAACCGC
GACACATTATGTATAGG TTATCA CCGAACAATTCA
CTGTAGACACAGTACTACAAAA AATGTAACAGTA
TGTTAACCTTCTAGAAGA AAGCATA CGGGAAACTATGCA...

Food flavors

Nestlé

FERRERO

Drugs

gsk

MSD

Pfizer

Human variation

Enzymes can be used to naturally produce food flavours and fragrances but also drugs!

Product Safety

head&shoulders
green apple
ADVANCED FORMULA
head&shoulders
conditioner
green apple

With Hao (BII) + ICE2 + IHPC +SIFBI

Jhoann

AI (CNN) prediction of mutation effects on enzyme function

PHARMA INNOVATION
PROGRAMME SINGAPORE (PIPS)

gsk
Pfizer
MSD

Kinetics

EVB Model

Binding

Docking scores

Binding free energy

r=0.9

Stability

Yasara + FoldX vs Rosetta

8 of 12 correct predictions of increased stability (incl. top 3)

Aggregation

Tango
Waltz

1 predicted mutation set increases solubility 2-fold

Allostery

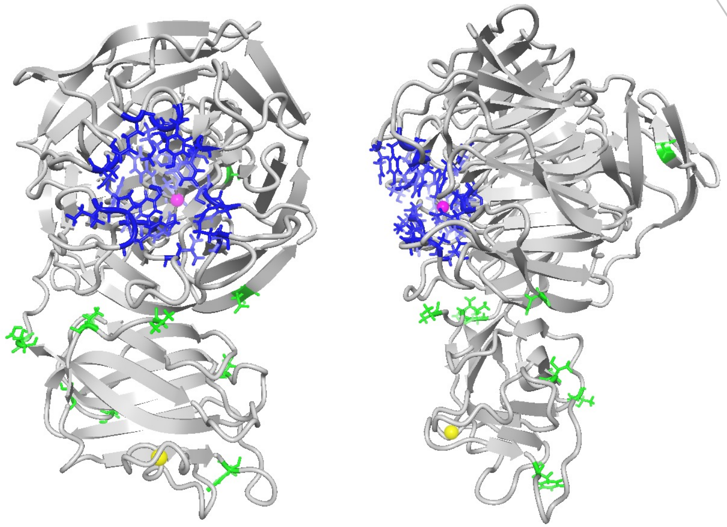
Allosigma

7 of 14 correct predictions of long range binding effect

Sequence Conservation

SIFT
Shannon entropy

Jump from 11% good sites to 67% good sites



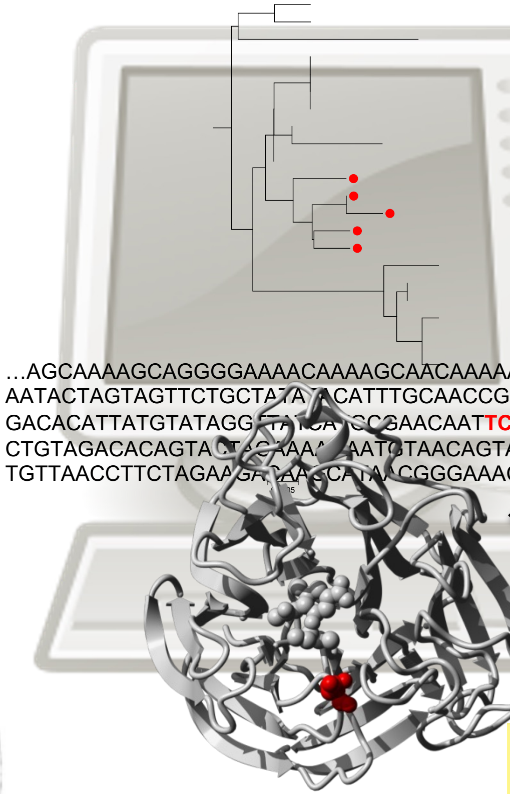
Restricted

Viruses

Computational Sequence and Structure Analysis

Food flavors

Drugs


…AGCAAAAGCAGGGGAAAACAAAAGCAACAAAAATGAAGGC
AATACTAGTAGTTCTGCTATA ACATTTGCAACCGCAAATGCA
GACACATTATGTATAGG TA CA CCGAACAATTCAACAGACA
CTGTAGACACAGTA AAA AATGTAACAGTAACAC CTC
TGTTAACCTTCTAGAAGA AA CATA CGGGAAACTAT

Human variation

Product safety

P&G

head & shoulders
ADVANCED FORMULA
head & shoulders
nditioner
en apple
IMPOSSIBLE
"WHAT THE !?"

Singapore Food Agency

Proteins can cause allergy if similar to other allergens.

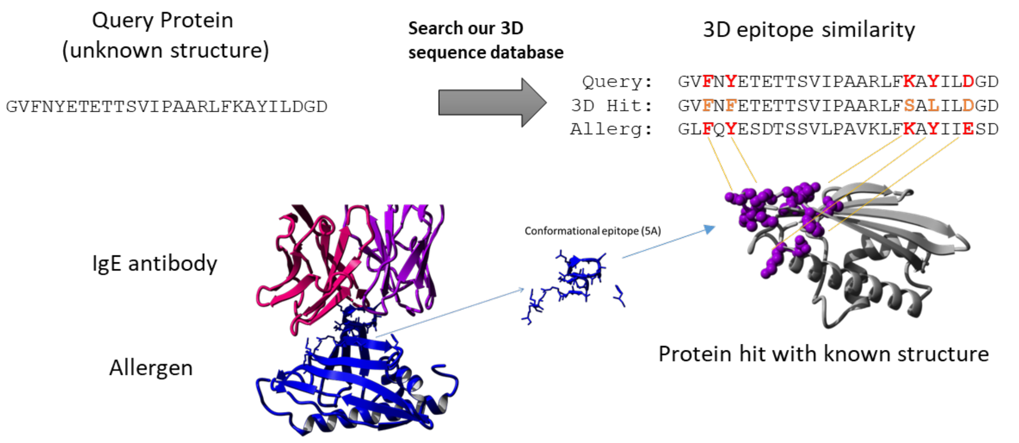
Computational prediction of protein allergenicity: from viruses to shampoo and food safety

Vachi, Minh

Path to new impacts:

- Started with flu antibody binding modelling expertise
- New industry application of similar method applied to antibody binding of allergens in consumer care products
- New area of application in food safety assessment

head & shoulders
IMPOSSIBLE
"WHAT THE !?"
CULTURED MEAT
LAB-GROWN CHICKEN



Query Protein
(unknown structure)

GVFNYETETTSVIPAARLFKAYILDGD

Search our 3D sequence database

3D epitope similarity

```
    Query:  GVFNYETETTSVIPAARLFKAYILDGD
  3D Hit:   GVFNFETETTSVIPAARLFSALILDGD
  Allerg:   GLFQYESDTSSVLPAVKLFKAYIIESD
```

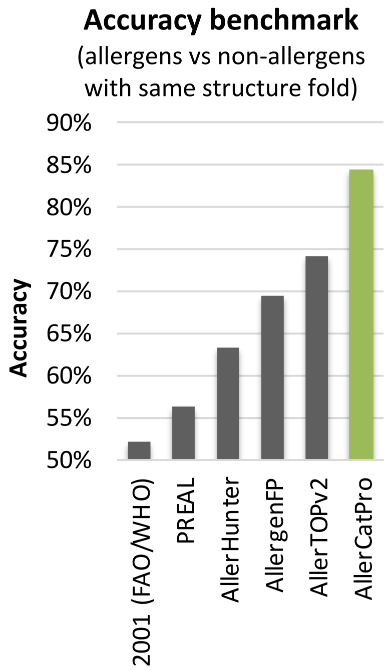
IgE antibody

Allergen

Conformational epitope (5A)

Protein hit with known structure

AllerCatPro
predicts if a protein is similar to known allergens



Accuracy benchmark
(allergens vs non-allergens with same structure fold)

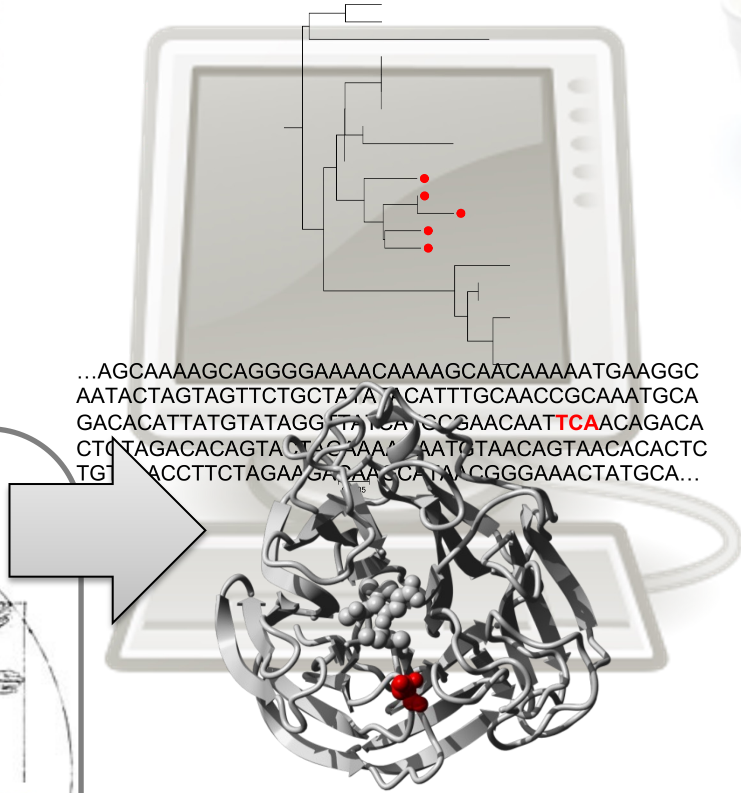
The impact:

- ✓ Industry:
  - ✓ 4 RCAs long-term work with P&G for co-development
  - ✓ active use by P&G in product development safety pipeline
  - ✓ Gluten extension work with The Coca Cola Company
- ✓ Supporting SG 30-by-30 goal by providing tool to assess risk of alternative/novel protein sources for food
  - ✓ Project with CSIRO/JCU on seafood allergy and risk of insect proteins as food source
  - ✓ Consulting for Singapore Food Agency for ongoing safety assessments in expert panel

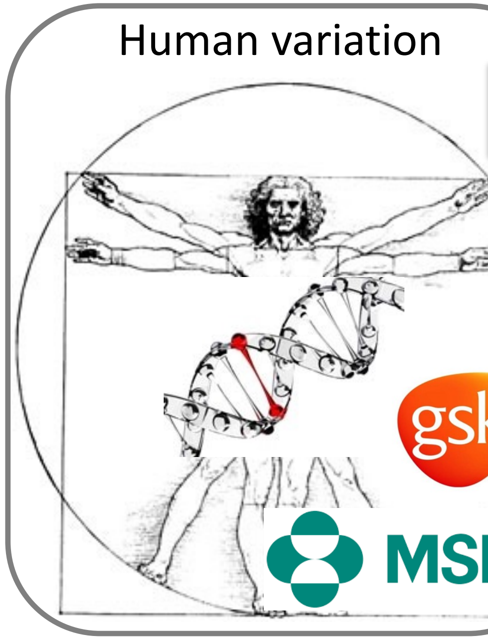
Computational Sequence and Structure Analysis

Viruses

Food flavors


…AGCAAAAGCAGGGGAAAACAAAAGCAACAAAAATGAAGGC
AATACTAGTAGTTCTGCTATATACATTTGCAACCGCAAATGCA
GACACATTATGTATAGGTTATCATGCGAACAATTCAACAGACA
CTGTAGACACAGTACTAGAAAAGAATGTAACAGTAACACACTC
TGTTAACCTTCTAGAAGACAAGCATAACGGGAAACTATGCA…

Human variation



GIS

Individual mutations can cause disease or affect drug efficacy

Product Safety

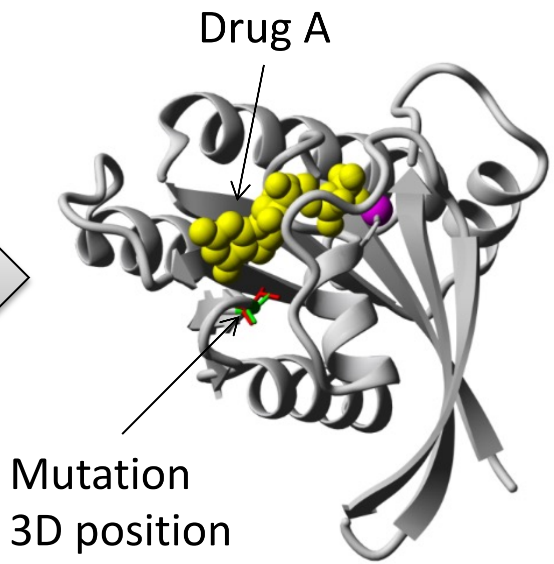
Bioinformatics Institute
BII
Genome Institute of Singapore
GIS

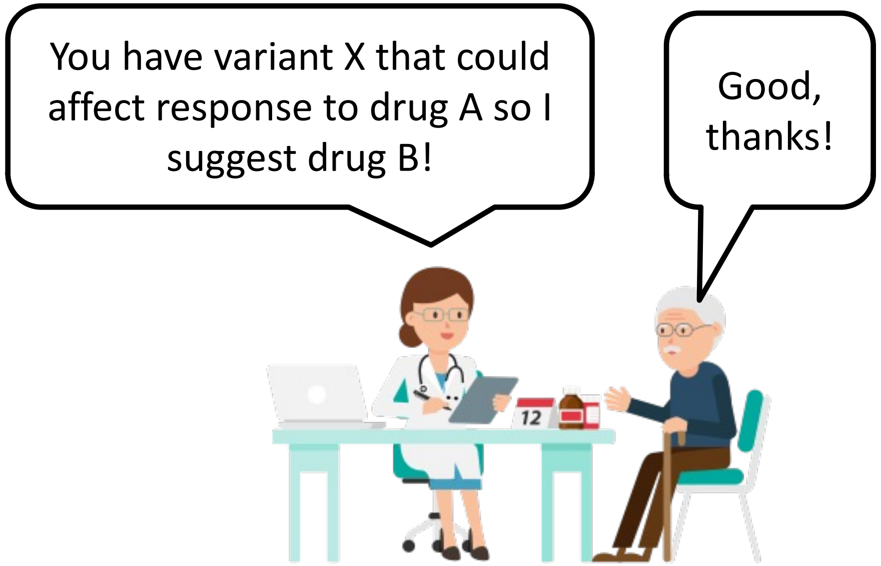
Ashar, Dimitar, Ken Chandra, Sebastian

SNPdrug3D database for Precision Medicine

...AGCAAAAGCAGGGGAAAACAAAAGCAA
CAAAAATGAAGGCAATACTAGTAGTTCTG
CTATATACATTTGCAACCGCAAATGCAGA
CACATTATGTATAGGTTATCATGCGAACAA
T**TCA**ACAGACACTGTAGACACAGTACTAG
AAAAGAATGTAACAGTAACACACTCTGTTA
ACCTTCTAGAAGACAAGCATAACGGGAAA
CTATGCA...

Sequence SNP/variant X


Drug A
Mutation 3D position

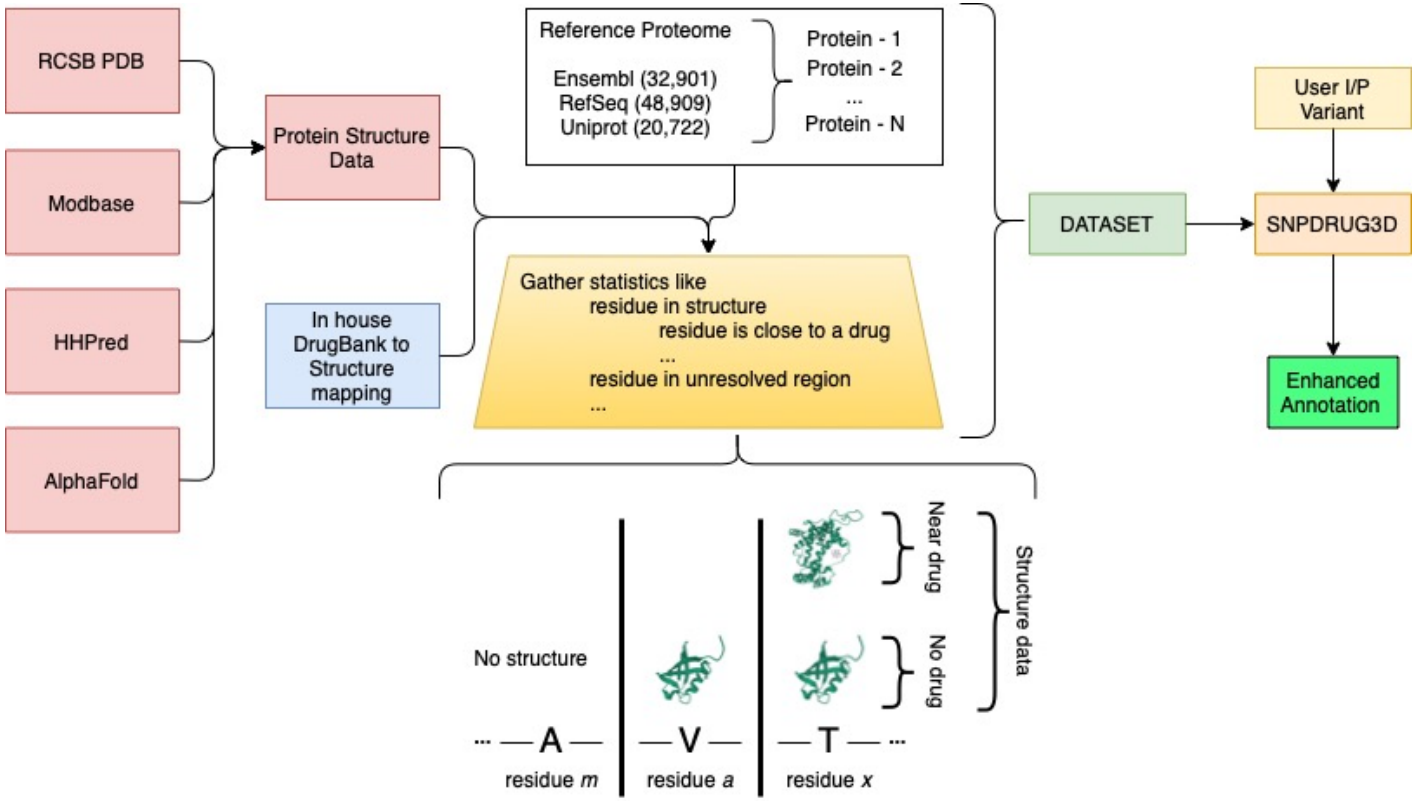

You have variant X that could affect response to drug A so I suggest drug B!
Good, thanks!
12

https://commons.wikimedia.org/wiki/File:Doctor_with_Patient_Cartoon.svg

User	Value
Scientist	SNP prioritization VUS annotation
Clinician scientist	SNP prioritization VUS annotation Known and new PGx SNPs
Clinician	Known PGx SNPs

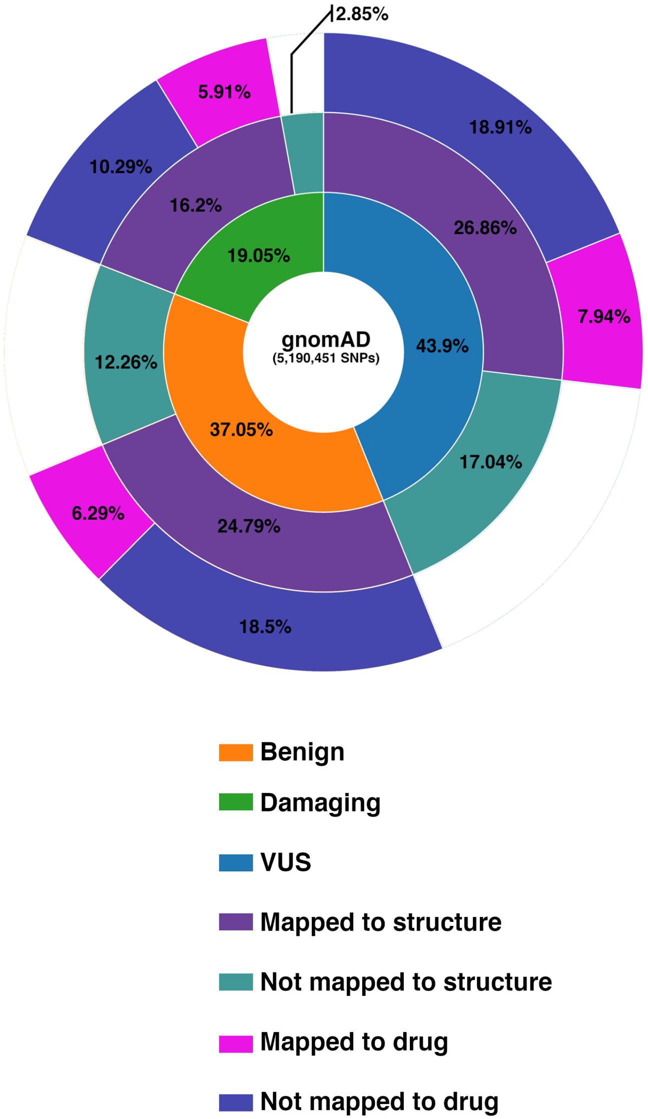
Ashar

SNPdrug3D database for Precision Medicine



RCSB PDB
Modbase
HHPred
AlphaFold
Protein Structure Data
In house DrugBank to Structure mapping
Reference Proteome
Ensembl (32,901)
RefSeq (48,909)
Uniprot (20,722)
Protein - 1
Protein - 2
...
Protein - N
Gather statistics like
residue in structure
residue is close to a drug
...
residue in unresolved region
...
DATASET
User I/P Variant
SNPDRUG3D
Enhanced Annotation

Near drug
No drug
Structure data
No structure
... — A — — V — — T — ...
residue *m*
residue *a*
residue *x*

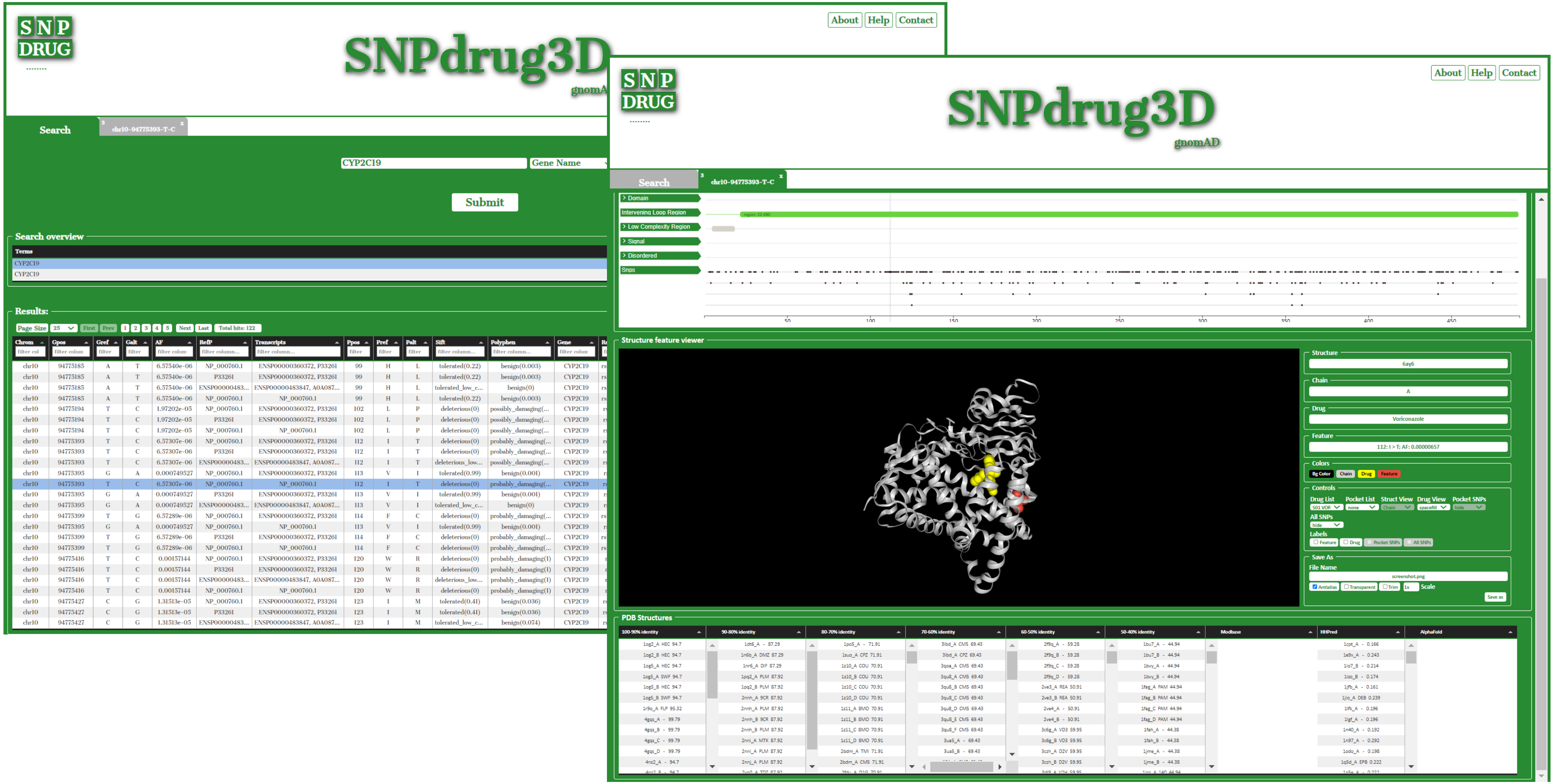


Dimitar

Bioinformatics Institute
BII

Genome Institute of Singapore
GIS

SNPdrug3D database for Precision Medicine



CYP2C19

Search overview

Terms
CYP2C19
CYP2C19

Results:

Page Size 25 · First · Prev · 1 2 3 4 5 · Next · Last · Total hits: 122

Chrom	Gpos	Gref	Galt	AF	RefP	Transcripts	Ppos	Pref	Palt	Sift	Polyphen	Gene
chr10	94775185	A	T	6.57540e-06	NP_000760.1	ENSP00000360372, P33261	99	H	L	tolerated(0.22)	benign(0.003)	CYP2C19
chr10	94775185	A	T	6.57540e-06	P33261	ENSP00000360372, P33261	99	H	L	tolerated(0.22)	benign(0.003)	CYP2C19
chr10	94775185	A	T	6.57540e-06	ENSP00000483...	ENSP00000483847, A0A087...	99	H	L	tolerated_low_c...	benign(0)	CYP2C19
chr10	94775185	A	T	6.57540e-06	NP_000760.1	NP_000760.1	99	H	L	tolerated(0.22)	benign(0.003)	CYP2C19
chr10	94775194	T	C	1.97202e-05	NP_000760.1	ENSP00000360372, P33261	102	L	P	deleterious(0)	possibly_damaging(...	CYP2C19
chr10	94775194	T	C	1.97202e-05	P33261	ENSP00000360372, P33261	102	L	P	deleterious(0)	possibly_damaging(...	CYP2C19
chr10	94775194	T	C	1.97202e-05	NP_000760.1	NP_000760.1	102	L	P	deleterious(0)	possibly_damaging(...	CYP2C19
chr10	94775393	T	C	6.57307e-06	NP_000760.1	ENSP00000360372, P33261	112	I	T	deleterious(0)	probably_damaging(...	CYP2C19
chr10	94775393	T	C	6.57307e-06	P33261	ENSP00000360372, P33261	112	I	T	deleterious(0)	probably_damaging(...	CYP2C19
chr10	94775393	T	C	6.57307e-06	ENSP00000483...	ENSP00000483847, A0A087...	112	I	T	deleterious_low...	possibly_damaging(...	CYP2C19
chr10	94775395	G	A	0.000749527	NP_000760.1	ENSP00000360372, P33261	113	V	I	tolerated(0.99)	benign(0.001)	CYP2C19
chr10	94775393	T	C	6.57307e-06	NP_000760.1	NP_000760.1	112	I	T	deleterious(0)	probably_damaging(...	CYP2C19
chr10	94775395	G	A	0.000749527	P33261	ENSP00000360372, P33261	113	V	I	tolerated(0.99)	benign(0.001)	CYP2C19
chr10	94775395	G	A	0.000749527	ENSP00000483...	ENSP00000483847, A0A087...	113	V	I	tolerated_low_c...	benign(0)	CYP2C19
chr10	94775399	T	G	6.57289e-06	NP_000760.1	ENSP00000360372, P33261	114	F	C	deleterious(0)	probably_damaging(...	CYP2C19
chr10	94775395	G	A	0.000749527	NP_000760.1	NP_000760.1	113	V	I	tolerated(0.99)	benign(0.001)	CYP2C19
chr10	94775399	T	G	6.57289e-06	P33261	ENSP00000360372, P33261	114	F	C	deleterious(0)	probably_damaging(...	CYP2C19
chr10	94775399	T	G	6.57289e-06	NP_000760.1	NP_000760.1	114	F	C	deleterious(0)	probably_damaging(...	CYP2C19
chr10	94775416	T	C	0.00157144	NP_000760.1	ENSP00000360372, P33261	120	W	R	deleterious(0)	probably_damaging(1)	CYP2C19
chr10	94775416	T	C	0.00157144	P33261	ENSP00000360372, P33261	120	W	R	deleterious(0)	probably_damaging(1)	CYP2C19
chr10	94775416	T	C	0.00157144	ENSP00000483...	ENSP00000483847, A0A087...	120	W	R	deleterious_low...	probably_damaging(1)	CYP2C19
chr10	94775416	T	C	0.00157144	NP_000760.1	NP_000760.1	120	W	R	deleterious(0)	probably_damaging(1)	CYP2C19
chr10	94775427	C	G	1.31513e-05	NP_000760.1	ENSP00000360372, P33261	123	I	M	tolerated(0.41)	benign(0.036)	CYP2C19
chr10	94775427	C	G	1.31513e-05	P33261	ENSP00000360372, P33261	123	I	M	tolerated(0.41)	benign(0.036)	CYP2C19
chr10	94775427	C	G	1.31513e-05	ENSP00000483...	ENSP00000483847, A0A087...	123	I	M	tolerated_low_c...	benign(0.074)	CYP2C19

Structure
6ay6

Chain
A

Drug
Voriconazole

Feature
112: I > T; AF: 0.00000657

Controls
Drug List: 501 VOR
Pocket List: none
Struct View: Chain
Drug View: spacefill
Pocket SNPs: hide
All SNPs: hide
Labels: ☐ Feature ☐ Drug ☐ Pocket SNPs ☐ All SNPs

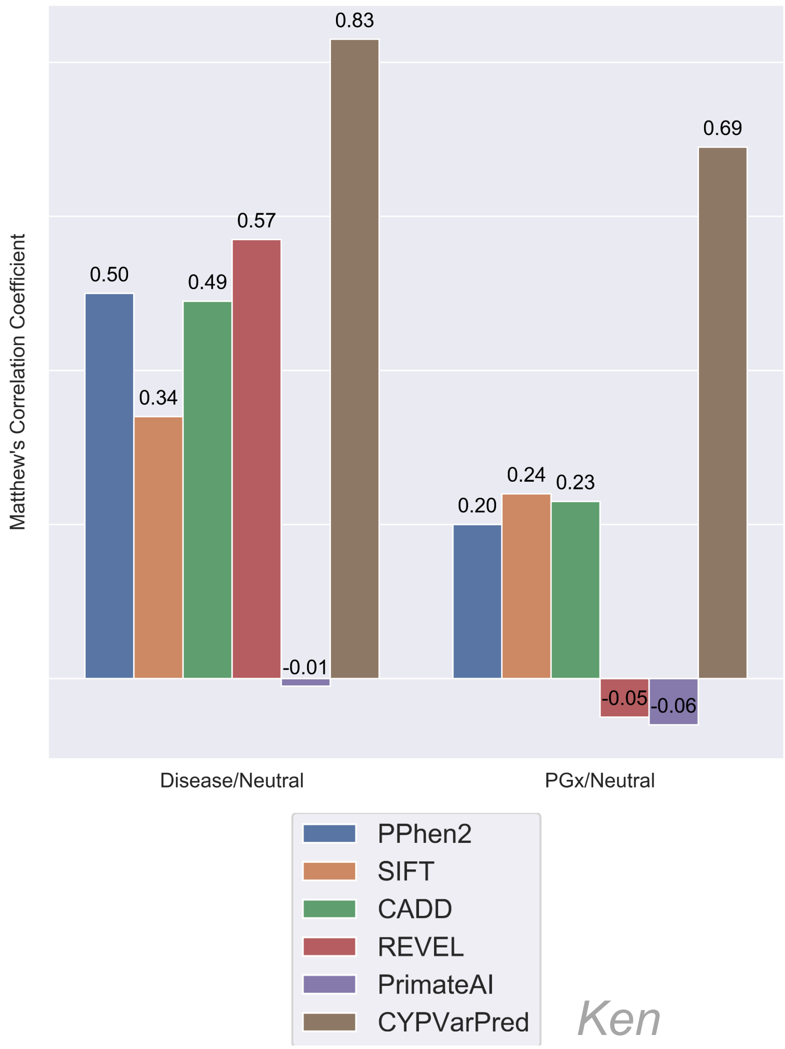
Save As
File Name: screenshot.png
☑ Antialias ☐ Transparent ☐ Trim 1x Scale

PDB Structures

100-90% identity	90-80% identity	80-70% identity	70-60% identity	60-50% identity	50-40% identity	Modbase	HHPred	AlphaFold
1og2_A HEC 94.7	1dt6_A - 87.29	1po5_A - 71.91	3ibd_A CM5 69.43	2f9q_A - 59.28	1bu7_A - 44.94		1cpt_A - 0.166	
1og2_B HEC 94.7	1n6b_A DMZ 87.29	1suo_A CPZ 71.91	3ibd_A CPZ 69.43	2f9q_B - 59.28	1bu7_B - 44.94		1e9x_A - 0.243	
1og5_A HEC 94.7	1nr6_A DIF 87.29	1z10_A COU 70.91	3qoa_A CM5 69.43	2f9q_C - 59.28	1bvy_A - 44.94		1io7_B - 0.214	
1og5_A SWF 94.7	1pq2_A PLM 87.92	1z10_B COU 70.91	3qu8_A CM5 69.43	2f9q_D - 59.28	1bvy_B - 44.94		1izo_B - 0.174	
1og5_B HEC 94.7	1pq2_B PLM 87.92	1z10_C COU 70.91	3qu8_B CM5 69.43	2ve3_A REA 50.91	1fag_A PAM 44.94		1jfb_A - 0.161	
1og5_B SWF 94.7	2nnh_A 9CR 87.92	1z10_D COU 70.91	3qu8_C CM5 69.43	2ve3_B REA 50.91	1fag_B PAM 44.94		1jio_A DEB 0.239	
1r9o_A FLP 95.32	2nnh_A PLM 87.92	1z11_A 8MO 70.91	3qu8_D CM5 69.43	2ve4_A - 50.91	1fag_C PAM 44.94		1lfk_A - 0.196	
4gqs_A - 99.79	2nnh_B 9CR 87.92	1z11_B 8MO 70.91	3qu8_E CM5 69.43	2ve4_B - 50.91	1fag_D PAM 44.94		1lgf_A - 0.196	
4gqs_B - 99.79	2nnh_B PLM 87.92	1z11_C 8MO 70.91	3qu8_F CM5 69.43	3c6g_A VD3 59.95	1fah_A - 44.38		1n40_A - 0.192	
4gqs_C - 99.79	2nni_A MTK 87.92	1z11_D 8MO 70.91	3ua5_A - 69.43	3c6g_B VD3 59.95	1fah_B - 44.38		1n97_A - 0.292	
4gqs_D - 99.79	2nni_A PLM 87.92	2bdm_A TMI 71.91	3ua5_B - 69.43	3czh_A D2V 59.95	1jme_A - 44.38		1odo_A - 0.198	
4nz2_A - 94.7	2nnj_A PLM 87.92	2bdm_A CM5 71.91		3czh_B D2V 59.95	1jme_B - 44.38		1q5d_A EPB 0.222	

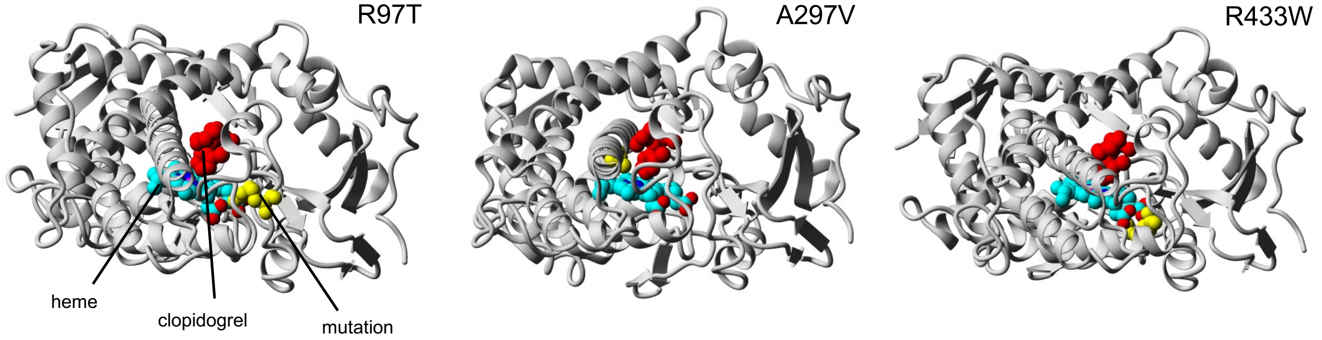
CREATING GROWTH, ENHANCING LIVES

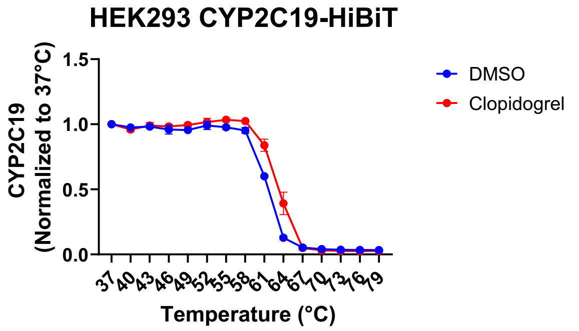
SNPdrug3D database for Precision Medicine

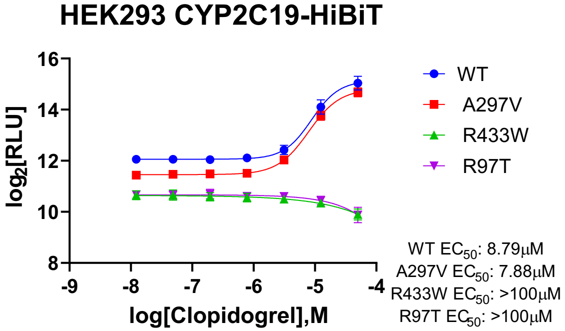


Ken

Drug metabolizing enzyme (CYP2C19)


R97T
A297V
R433W
heme
clopidogrel
mutation





Chris Tan team

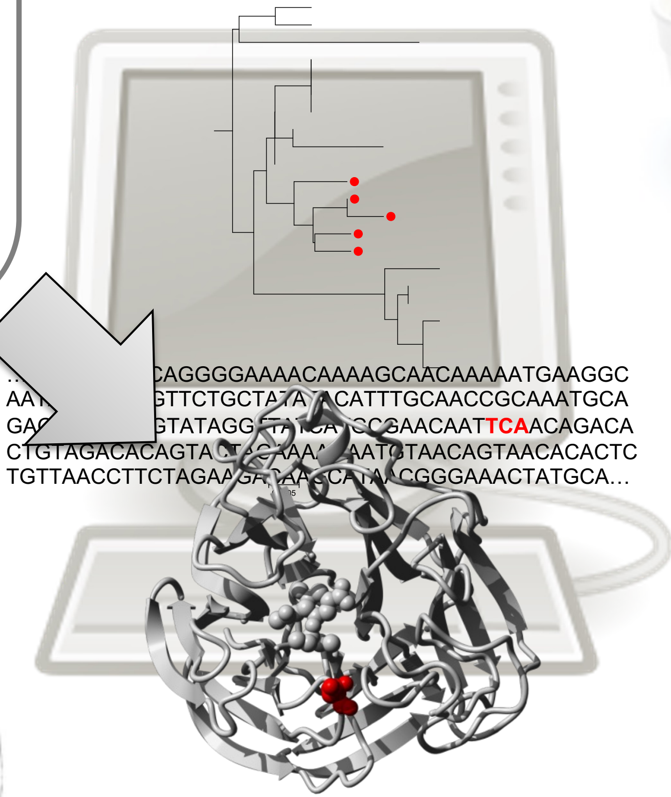
GIS

Viruses

GISAID
World Health Organization

Restricted

Computational Sequence and Structure Analysis


...CAGGGGAAAACAAAAGCAACAAAAATGAAGGC
AAT GTTCTGCTATA ACATTTGCAACCGCAAATGCA
GAC GTATAGG TA CA CCGAACAATTCAACAGACA
CTGTAGACACAGTA AAA AATGTAACAGTAACACACTC
TGTTAACCTTCTAGAAGA AA CA A CGGGAAACTATGCA...

Food flavors

Drugs

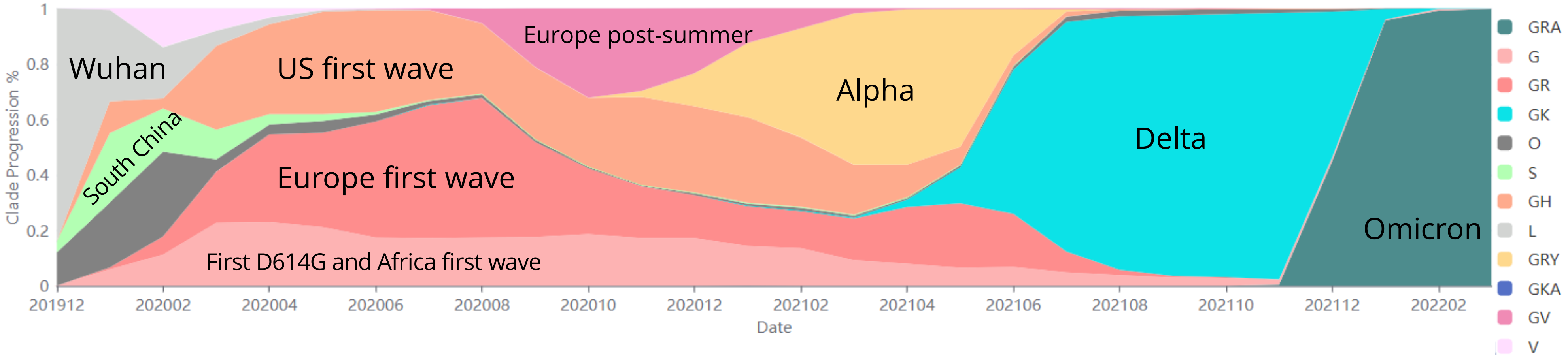
Human variation

Long history of flu work but COVID put us into the spotlight

Product Safety

head & shoulders
head & shoulders
ADVANCED FORMULA
conditioner
green apple
green apple

POWERING DISCOVERIES



Our BII team is part of the history of this pandemic and a pillar of GISAID and in Singapore’s COVID response

GISAID

Curation: Shruti, Yi Hong, Meera *Annotation, tools, reports:* Raphael, Yani, Joses, Sandy, Dimitar, Ken

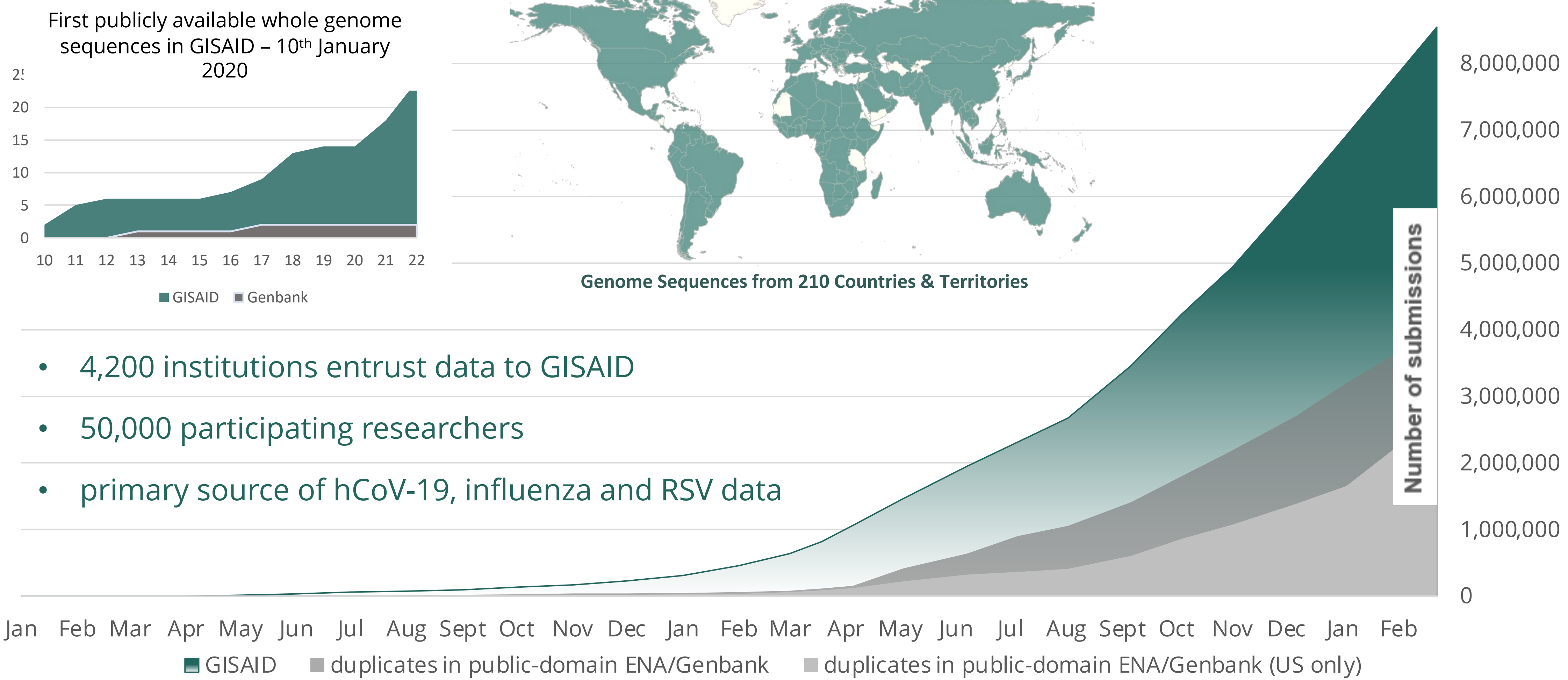
Shruti, Yi Hong, Meera

Real-time genomic data sharing during the COVID-19 Pandemic

GISAID's transparent sharing mechanism is favored over anonymous sharing via public-domain

enabled first vaccines (Polack et al *N Engl J Med* 2020)
enabled first diagnostic tests (Bohn et al *Clin Chem Lab Med* 2020)

In SG: Contributed to Fortitude and Resolute kits



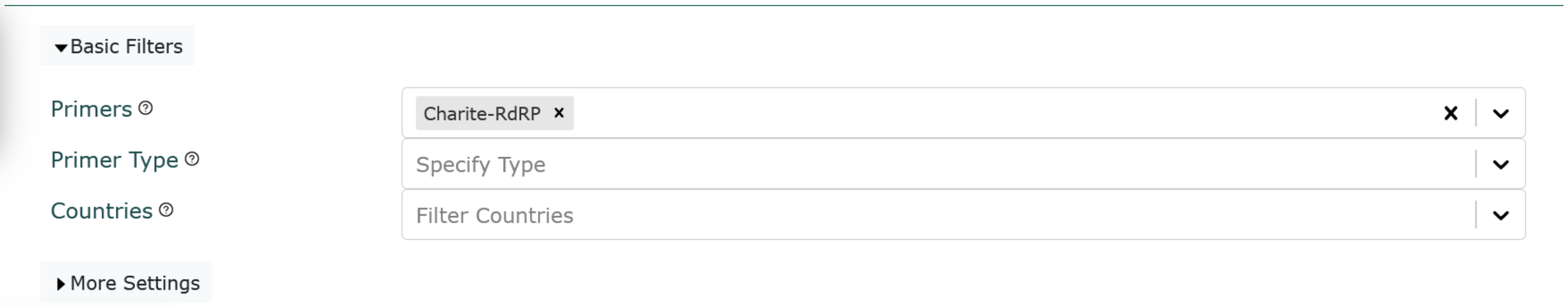
- 4,200 institutions entrust data to GISAID
- 50,000 participating researchers
- primary source of hCoV-19, influenza and RSV data

BII with help from GIS has started and complemented the global team that curated all genomes on GISAID since day 1 and every single genome goes through our servers for error checking and mutation annotation

POWERING DISCOVERIES

Our tool enables real-time primer tracking for Dx

PrimerChecker



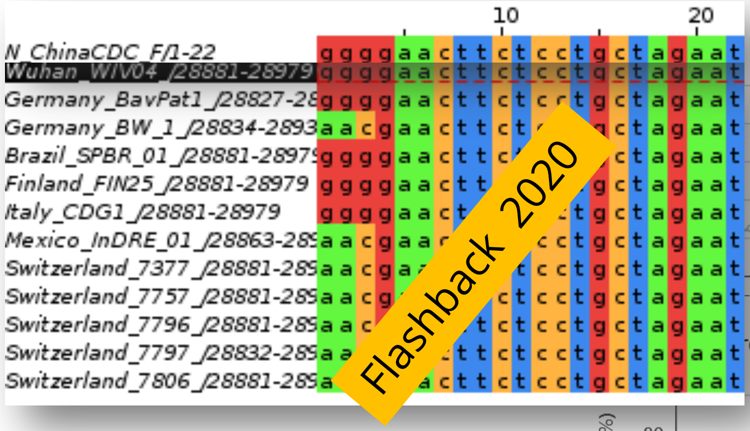
▾ Basic Filters

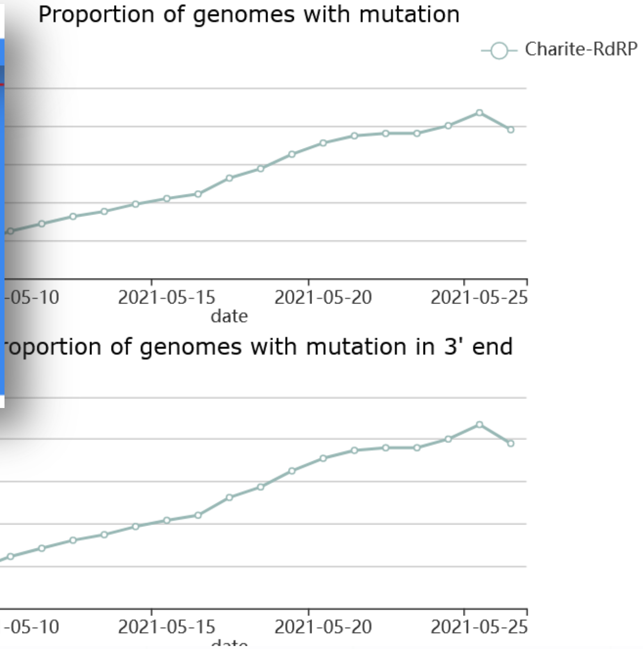
Primers ⓘ: Charite-RdRP ×

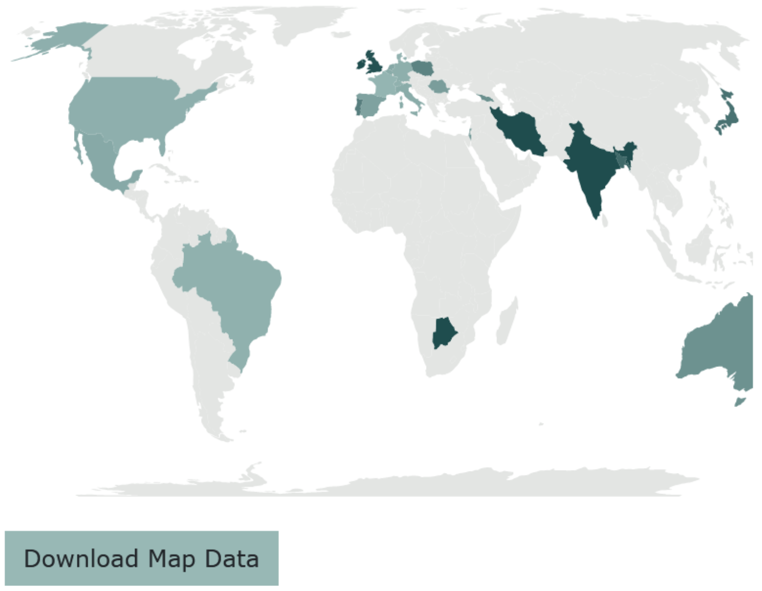
Primer Type ⓘ: Specify Type

Countries ⓘ: Filter Countries

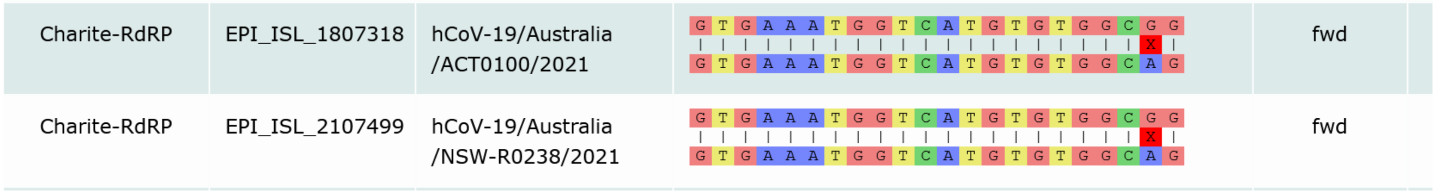
▸ More Settings


10
20
N_ChinaCDC_F/1-22
Wuhan_WIV04_/28881-28979
Germany_BavPat1_/28827-28
Germany_BW_1_/28834-2893
Brazil_SPBR_01_/28881-2897
Finland_FIN25_/28881-28979
Italy_CDG1_/28881-28979
Mexico_InDRE_01_/28863-28
Switzerland_7377_/28881-28
Switzerland_7757_/28881-28
Switzerland_7796_/28881-28
Switzerland_7797_/28832-28
Switzerland_7806_/28881-28
Flashback 2020




Download Map Data

E.g. the team alerted for mismatches in China CDC N and Charite RdRP primers when they emerged



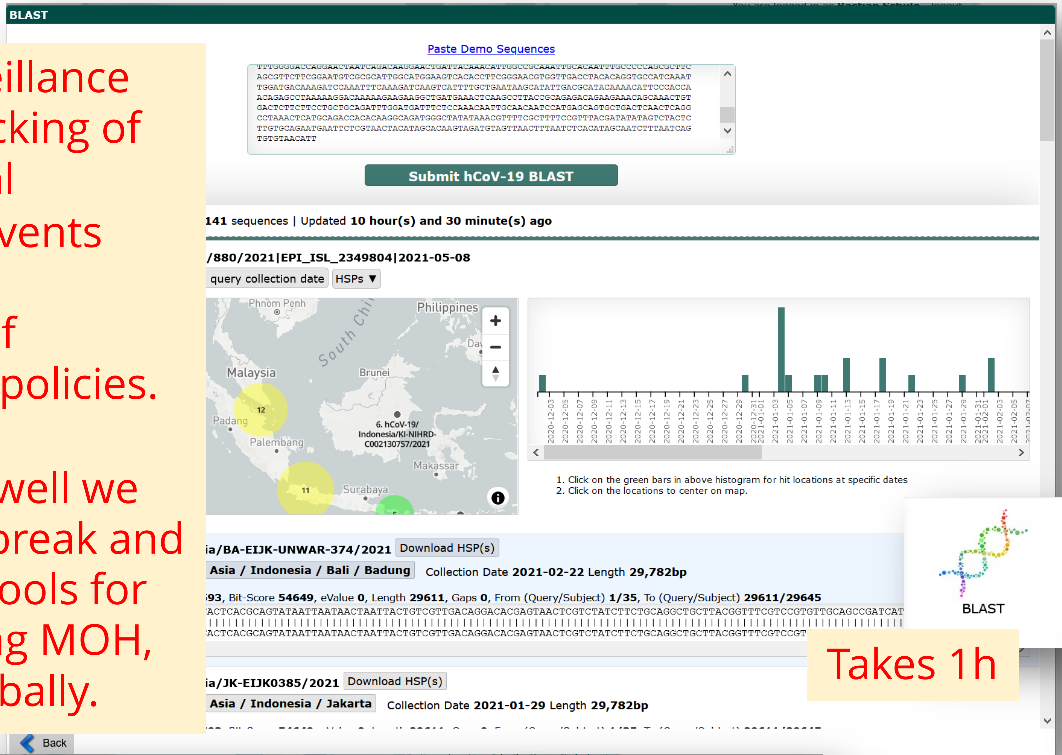
Charite-RdRP	EPI_ISL_1807318	hCoV-19/Australia/ACT0100/2021	G T G A A A T G G T C A T G T G T G G C G G | | | | | | | | | | | | | | | | | | | | X | G T G A A A T G G T C A T G T G T G G C A G	fwd
Charite-RdRP	EPI_ISL_2107499	hCoV-19/Australia/NSW-R0238/2021	G T G A A A T G G T C A T G T G T G G C G G | | | | | | | | | | | | | | | | | | | | X | G T G A A A T G G T C A T G T G T G G C A G	fwd

Sandy

Expertise and tools supporting contact tracing

Genomic surveillance also allows tracking of global and local transmission events which informs effectiveness of measures and policies.

It can tell how well we control an outbreak and BII is building tools for that. Supporting MOH, ASEAN and globally.



BLAST

Paste Demo Sequences

Submit hCoV-19 BLAST

141 sequences | Updated **10 hour(s) and 30 minute(s) ago**

/880/2021|EPI_ISL_2349804|2021-05-08

query collection date HSPs ▼

Phnom Penh
Philippines
South Chi
Malaysia
Brunei
Padang
Palembang
Makassar
Surabaya
12
11
6. hCoV-19/ Indonesia/KI-NIHRD-C002130757/2021
2020-12-03
2021-02-05

1. Click on the green bars in above histogram for hit locations at specific dates
2. Click on the locations to center on map.

ia/BA-EIJK-UNWAR-374/2021 Download HSP(s)

Asia / Indonesia / Bali / Badung Collection Date **2021-02-22** Length **29,782bp**

93, Bit-Score **54649**, eValue **0**, Length **29611**, Gaps **0**, From (Query/Subject) **1/35**, To (Query/Subject) **29611/29645**

```
ACTCACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGTGTTGCAGCCGATCAT
ACTCACGCAGTATAATTAATAACTAATTACTGTCGTTGACAGGACACGAGTAACTCGTCTATCTTCTGCAGGCTGCTTACGGTTTCGTCCGT
```

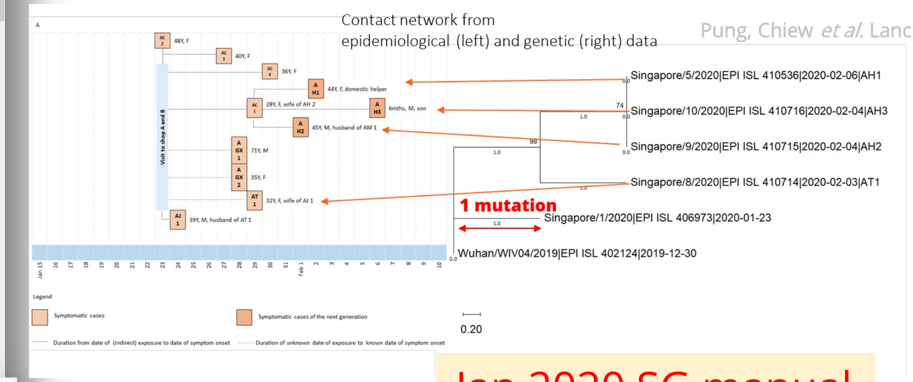
ia/JK-EIJK0385/2021 Download HSP(s)

Asia / Indonesia / Jakarta Collection Date **2021-01-29** Length **29,782bp**

Back

BLAST

Takes 1h

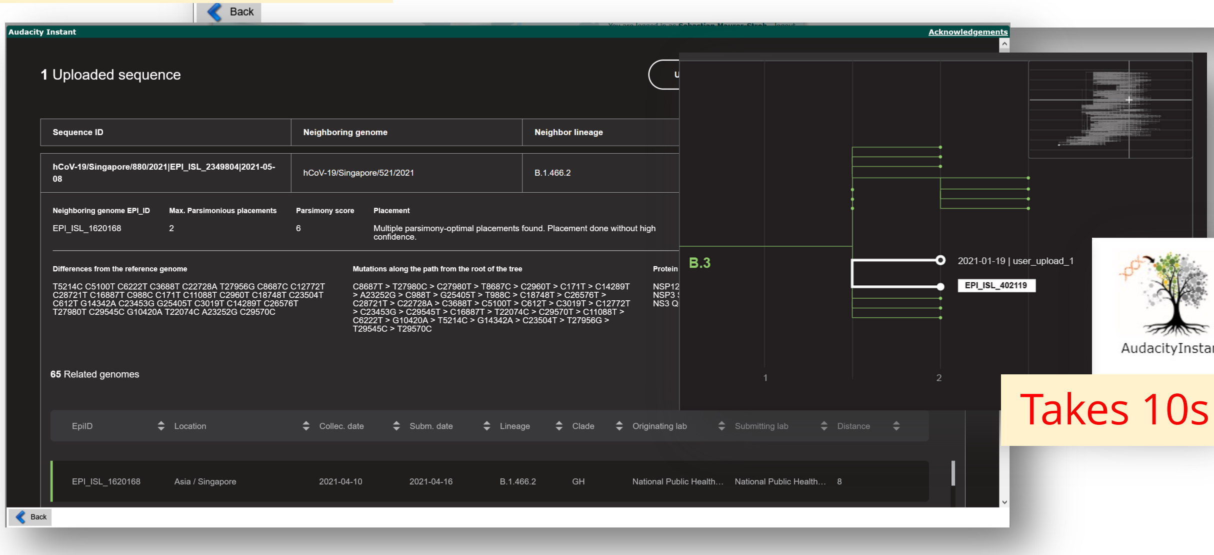


Contact network from epidemiological (left) and genetic (right) data

Pung, Chiew *et al.* Lancet

A
AC 2 48Y, F
AC 3 40Y, F
AC 4 36Y, F
A H1 44Y, F, domestic helper
AC 1 28Y, F, wife of AH 2
A H3 6mths, M, son
A H2 45Y, M, husband of AM 1
A GX 1 71Y, M
A GX 2 35Y, F
AT 1 32Y, F, wife of AJ 1
AJ 1 39Y, M, husband of AT 1
Visit to shop A and B
Jan 15
Feb 1
Legend
Symptomatic cases
Symptomatic cases of the next generation
Duration from date of (indirect) exposure to date of symptom onset
Duration of unknown date of exposure to known date of symptom onset
Singapore/5/2020|EPI ISL 410536|2020-02-06|AH1
Singapore/10/2020|EPI ISL 410716|2020-02-04|AH3
Singapore/9/2020|EPI ISL 410715|2020-02-04|AH2
Singapore/8/2020|EPI ISL 410714|2020-02-03|AT1
Singapore/1/2020|EPI ISL 406973|2020-01-23
Wuhan/WIV04/2019|EPI ISL 402124|2019-12-30
1 mutation
74
98
1.0
0.0
0.20

Jan 2020 SG manual



Audacity Instant

Acknowledgements

1 Uploaded sequence

Sequence ID	Neighboring genome	Neighbor lineage
hCoV-19/Singapore/880/2021|EPI_ISL_2349804|2021-05-08	hCoV-19/Singapore/521/2021	B.1.466.2

Neighboring genome EPI_ID	Max. Parsimonious placements	Parsimony score	Placement
EPI_ISL_1620168	2	6	Multiple parsimony-optimal placements found. Placement done without high confidence.

Differences from the reference genome

T5214C C5100T C6222T C3688T C22728A T27956G C8687C C12772T C28721T C16887T C988C C171T C11088T C2960T C18748T C23504T C612T G14342A C23453G G25405T C3019T C14289T C26576T T27980T C29545C G10420A T22074C A23252G C29570C

Mutations along the path from the root of the tree

C8687T > T27980C > C27980T > T8687C > C2960T > C171T > C14289T > A23252G > C988T > G25405T > T988C > C18748T > C26576T > C28721T > C22728A > C3688T > C5100T > C612T > C3019T > C12772T > C23453G > C29545T > C16887T > T22074C > C29570T > C11088T > C6222T > G10420A > T5214C > G14342A > C23504T > T27956G > T29545C > T29570C

Protein

NSP12
NSP3
NS3 Q

65 Related genomes

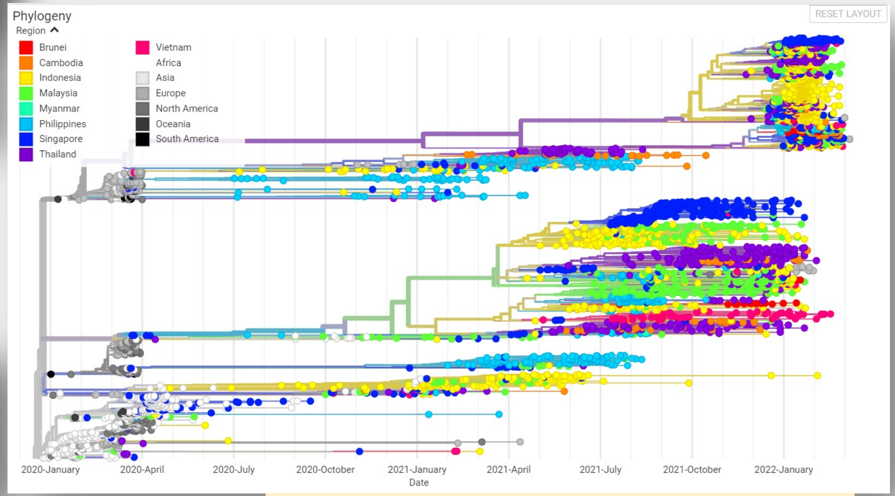
EpiID	Location	Collec. date	Subm. date	Lineage	Clade	Originating lab	Submitting lab	Distance
EPI_ISL_1620168	Asia / Singapore	2021-04-10	2021-04-16	B.1.466.2	GH	National Public Health…	National Public Health…	8

Back

B.3
2021-01-19 | user_upload_1
EPI_ISL_402119
1
2

AudacityInstant

Takes 10s

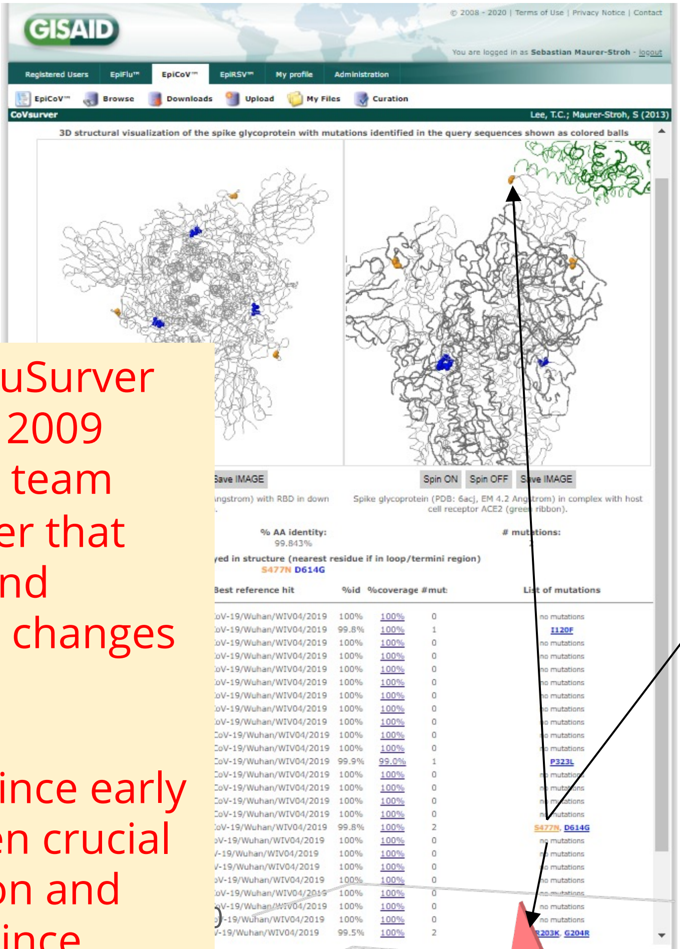


Jan 2022 ASEAN automated

POWERING DISCOVERIES

CoVsurver real-time surveillance for mutations that can affect vaccines etc.

Raphael



CoVsurver — Lee, T.C.; Maurer-Stroh, S (2013)

3D structural visualization of the spike glycoprotein with mutations identified in the query sequences shown as colored balls

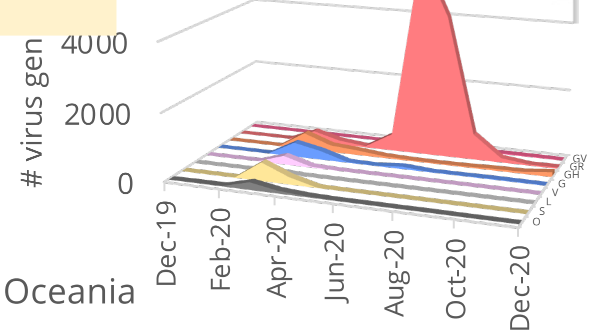
Spike glycoprotein (PDB: 6acj, EM 4.2 Angstrom) in complex with host cell receptor ACE2 (green ribbon).

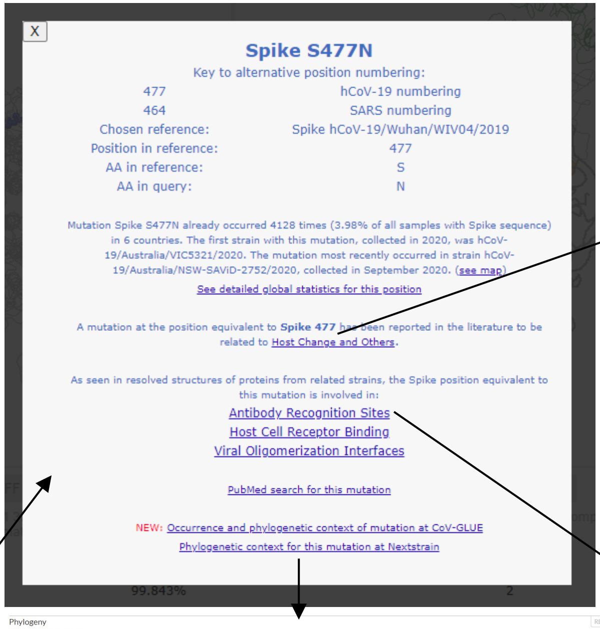
% AA identity: 99.843%

S477N D614G

Building on the FluSurver success since the 2009 pandemic, the BII team built the CoVsurver that can help to find and interpret relevant changes in the virus.

The tool is used since early 2020 and has been crucial for GISAID curation and annotation ever since.





Spike S477N

Key to alternative position numbering:

477	hCoV-19 numbering
464	SARS numbering
Chosen reference:	Spike hCoV-19/Wuhan/WIV04/2019
Position in reference:	477
AA in reference:	S
AA in query:	N

Mutation Spike S477N already occurred 4128 times (3.98% of all samples with Spike sequence) in 6 countries. The first strain with this mutation, collected in 2020, was hCoV-19/Australia/VIC5321/2020. The mutation most recently occurred in strain hCoV-19/Australia/NSW-SAViD-2752/2020, collected in September 2020. (see map)

See detailed global statistics for this position

A mutation at the position equivalent to **Spike 477** has been reported in the literature to be related to Host Change and Others.

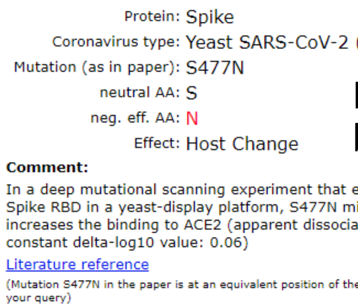
As seen in resolved structures of proteins from related strains, the Spike position equivalent to this mutation is involved in:

Antibody Recognition Sites
Host Cell Receptor Binding
Viral Oligomerization Interfaces

PubMed search for this mutation

NEW: Occurrence and phylogenetic context of mutation at CoV-GLUE

Phylogenetic context for this mutation at Nextstrain



Protein:	Spike
Coronavirus type:	Yeast SARS-CoV-2 (2019)
Mutation (as in paper):	S477N
neutral AA:	S
neg. eff. AA:	N
Effect:	Host Change

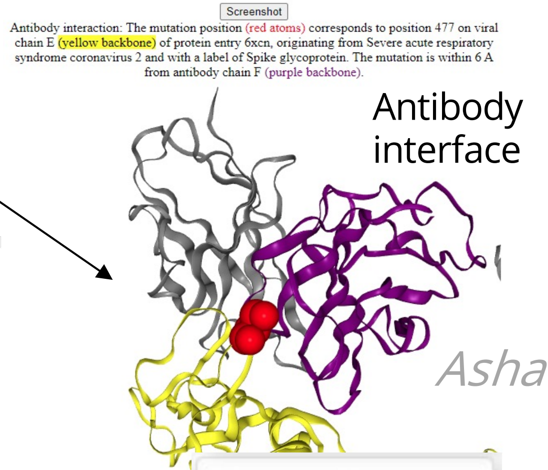
Comment:

In a deep mutational scanning experiment that expresses Spike RBD in a yeast-display platform, S477N mildly increases the binding to ACE2 (apparent dissociation constant delta-log10 value: 0.06)

Literature reference

(Mutation S477N in the paper is at an equivalent position of the mutation in your query)

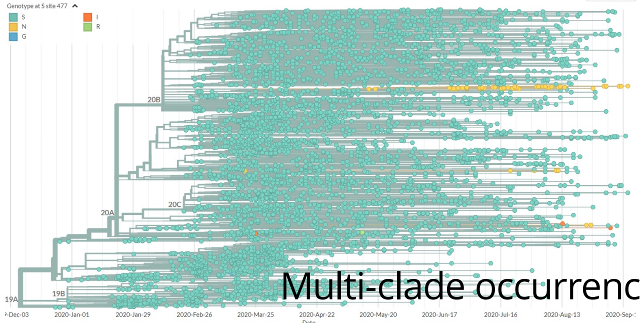
Receptor binding



Antibody interaction: The mutation position (red atoms) corresponds to position 477 on viral chain E (yellow backbone) of protein entry 6xcn, originating from Severe acute respiratory syndrome coronavirus 2 and with a label of Spike glycoprotein. The mutation is within 6 A from antibody chain F (purple backbone).

Antibody interface

Ashar



Multi-clade occurrence

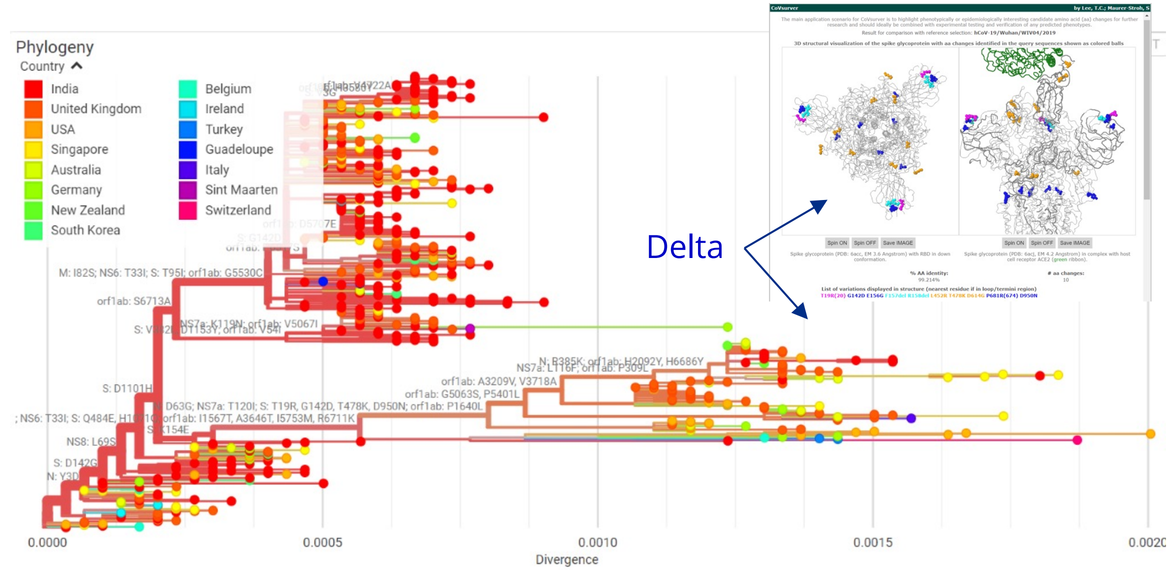
Spike glycoprotein mutation surveillance

1,680 variant phenotype annotations from literature plus 3,826 3D interactions

POWERING DISCOVERIES

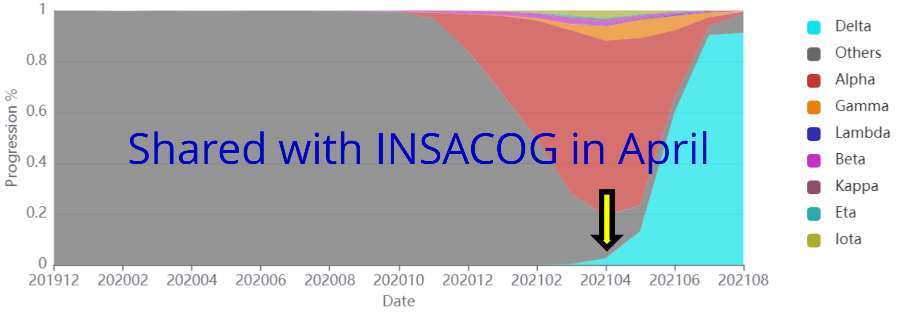
POWERING DISCOVERIES

Example for early phase of Delta and Omicron


Phylogeny
Country
India
United Kingdom
USA
Singapore
Australia
Germany
New Zealand
South Korea
Belgium
Ireland
Turkey
Guadeloupe
Italy
Sint Maarten
Switzerland
M: I82S; NS6: T33I; S: T95I; orf1ab: G5530C
orf1ab: S6713A
S: D1101H
; NS6: T33I; S: Q484E, H1101
NS8: L69S
S: D142G
N: Y3D
N: R385K; orf1ab: H2092Y, H6686Y
NS7a: L116F; orf1ab: P309L
orf1ab: A3209V, V3718A
orf1ab: G5063S, P5401L
D63G; NS7a: T120I; S: T19R, G142D, T478K, D950N; orf1ab: P1640L
orf1ab: I1567T, A3646T, I5753M, R6711K
S: E154E
0.0000
0.0005
0.0010
0.0015
0.0020
Divergence

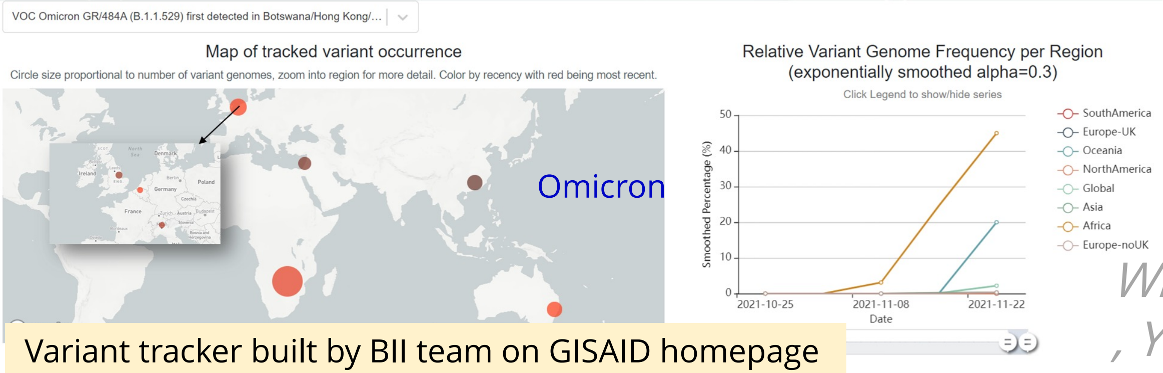
Delta

Flashback Apr 2021



Shared with INSACOG in April

Our team saw relevant changes in strains in India and alerted INSACOG and SG research teams before the strain exploded (it later got named Delta)


VOC Omicron GR/484A (B.1.1.529) first detected in Botswana/Hong Kong/...
Map of tracked variant occurrence
Circle size proportional to number of variant genomes, zoom into region for more detail. Color by recency with red being most recent.
Relative Variant Genome Frequency per Region
(exponentially smoothed alpha=0.3)
Click Legend to show/hide series
Smoothed Percentage (%)
50
40
30
20
10
0
2021-10-25
2021-11-08
2021-11-22
Date
SouthAmerica
Europe-UK
Oceania
NorthAmerica
Global
Asia
Africa
Europe-noUK

Omicron

Variant tracker built by BII team on GISAID homepage

Most recent reported occurrences in different countries

Country	Strain Name	Collection Date
Australia	hCoV-19/Australia/NSW-RPAH-1933/2021 ⓘ	2021-11-27
Belgium	hCoV-19/Belgium/rega-20174/2021 ⓘ	2021-11-24
Botswana	hCoV-19/Botswana/R43B70_BHP_4021000195/2021 ⓘ	2021-11-23
South Africa	hCoV-19/South Africa/CERI-KRISP-K032284/2021 ⓘ	2021-11-23
Italy	hCoV-19/Italy/LOM-Sacco_Var_T96348/2021 ⓘ	2021-11-22
United Kingdom	hCoV-19/England/MILK-2B67570/2021 ⓘ	2021-11-21
Israel	hCoV-19/Israel/SMC-7022800/2021 ⓘ	2021-11-20
Hong Kong	hCoV-19/Hong Kong/VM21045145/2021 ⓘ	2021-11-18

Country Submission Count

Country	Total #GR/484A (B.1.1.529)	#GR/484A (B.1.1.529) in past 4 weeks	%GR/484A (B.1.1.529) in past 4 weeks
South Africa	99	99	69.7
Botswana	19	19	18.6
Australia	2	2	0.1
Hong Kong	2	2	8.3
Italy	1	1	0.0
United Kingdom	1	1	0.0
Belgium	1	1	0.1
Israel	1	1	0.1

Winston, Yani, Raphael

Flashback Nov 2021

South African researchers worked with GISAID and BII to characterize and track Omicron

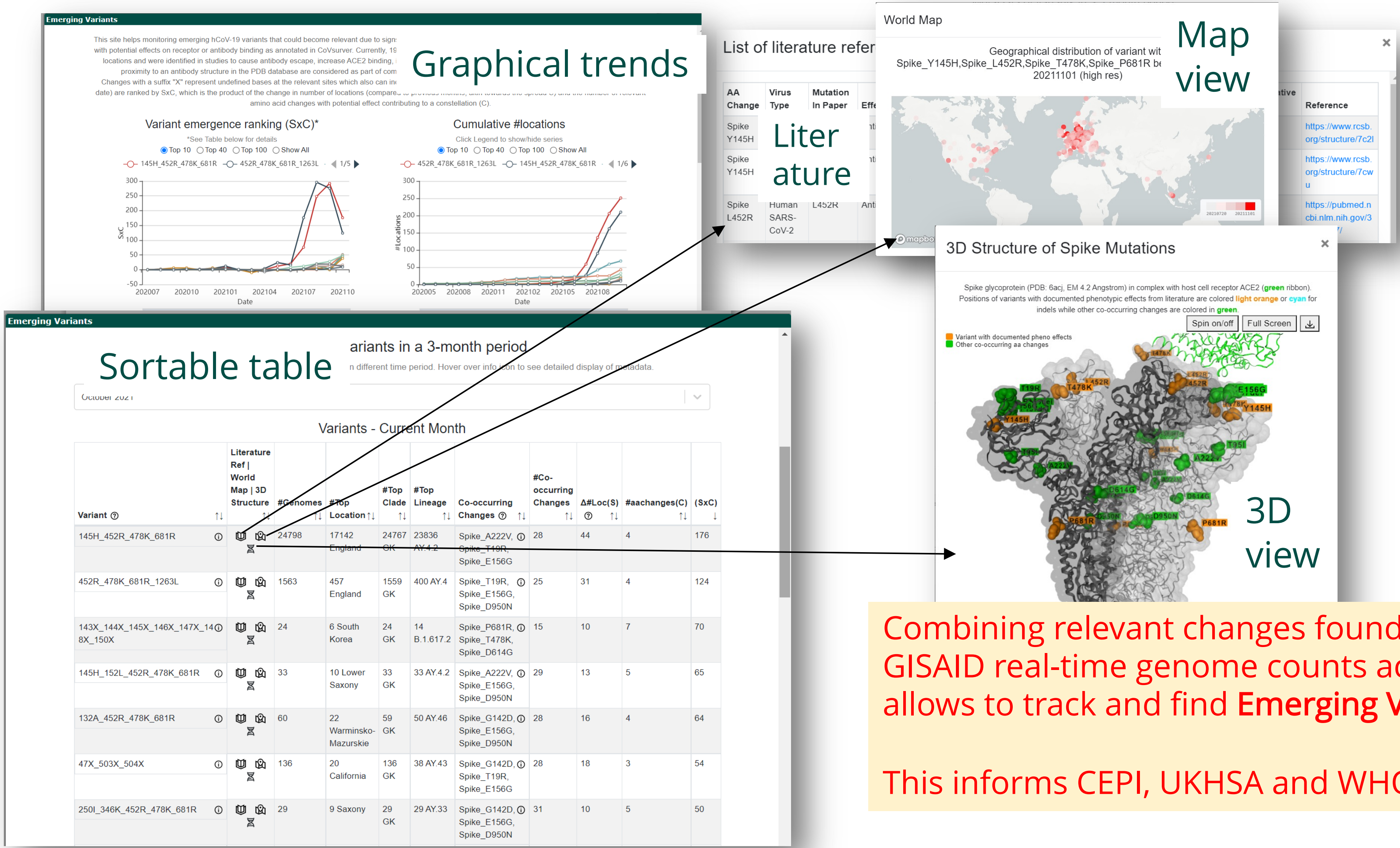
BII was also part of WHO taskforce for the variant naming scheme.

Supported by BII/GIS, A*STAR Singapore

Live dashboard for choosing which emerging variants to investigate

Raphael, Yani, Joses

- Fast spread (increase in observed locations)
- Number of relevant spike changes (with experimentally known effect or known Ab site)



Graphical trends

Emerging Variants

This site helps monitoring emerging hCoV-19 variants that could become relevant due to sign
with potential effects on receptor or antibody binding as annotated in CoVsurver. Currently, 19
locations and were identified in studies to cause antibody escape, increase ACE2 binding,
proximity to an antibody structure in the PDB database are considered as part of com
Changes with a suffix "X" represent undefined bases at the relevant sites which also can in
date) are ranked by SxC, which is the product of the change in number of locations (compare
amino acid changes with potential effect contributing to a constellation (C).

Variant emergence ranking (SxC)*
*See Table below for details
○ Top 10 ○ Top 40 ○ Top 100 ○ Show All
145H_452R_478K_681R
452R_478K_681R_1263L
1/5
SxC
Date
Cumulative #locations
Click Legend to show/hide series
○ Top 10 ○ Top 40 ○ Top 100 ○ Show All
452R_478K_681R_1263L
145H_452R_478K_681R
1/6
#Locations
Date
List of literature refer
AA Change
Virus Type
Mutation In Paper
Reference
Spike Y145H
Spike L452R
Human SARS-CoV-2
L452R
Liter ature
World Map
Geographical distribution of variant wit
Spike_Y145H,Spike_L452R,Spike_T478K,Spike_P681R be
20211101 (high res)
20210720
20211101
Map view
3D Structure of Spike Mutations
Spin on/off
Full Screen
Variant with documented pheno effects
Other co-occurring aa changes
T19R
T478K
L452R
E156G
Y145H
A222V
D614G
D950N
P681R
3D view

Spike glycoprotein (PDB: 6acj, EM 4.2 Angstrom) in complex with host cell receptor ACE2 (green ribbon). Positions of variants with documented phenotypic effects from literature are colored light orange or cyan for indels while other co-occurring changes are colored in green.

Sortable table

Emerging Variants

ariants in a 3-month period

n different time period. Hover over info icon to see detailed display of metadata.

October 2021

Variants - Current Month

Variant	Literature Ref | World Map | 3D Structure	#Genomes	#Top Location	#Top Clade	#Top Lineage	Co-occurring Changes	#Co-occurring Changes	Δ#Loc(S)	#aachanges(C)	(SxC)
145H_452R_478K_681R		24798	17142 England	24767 GK	23836 AY.4.2	Spike_A222V, Spike_T19R, Spike_E156G	28	44	4	176
452R_478K_681R_1263L		1563	457 England	1559 GK	400 AY.4	Spike_T19R, Spike_E156G, Spike_D950N	25	31	4	124
143X_144X_145X_146X_147X_148X_150X		24	6 South Korea	24 GK	14 B.1.617.2	Spike_P681R, Spike_T478K, Spike_D614G	15	10	7	70
145H_152L_452R_478K_681R		33	10 Lower Saxony	33 GK	33 AY.4.2	Spike_A222V, Spike_E156G, Spike_D950N	29	13	5	65
132A_452R_478K_681R		60	22 Warminsko-Mazurskie	59 GK	50 AY.46	Spike_G142D, Spike_E156G, Spike_D950N	28	16	4	64
47X_503X_504X		136	20 California	136 GK	38 AY.43	Spike_G142D, Spike_T19R, Spike_E156G	28	18	3	54
250I_346K_452R_478K_681R		29	9 Saxony	29 GK	29 AY.33	Spike_G142D, Spike_E156G, Spike_D950N	31	10	5	50

Emerging Variants

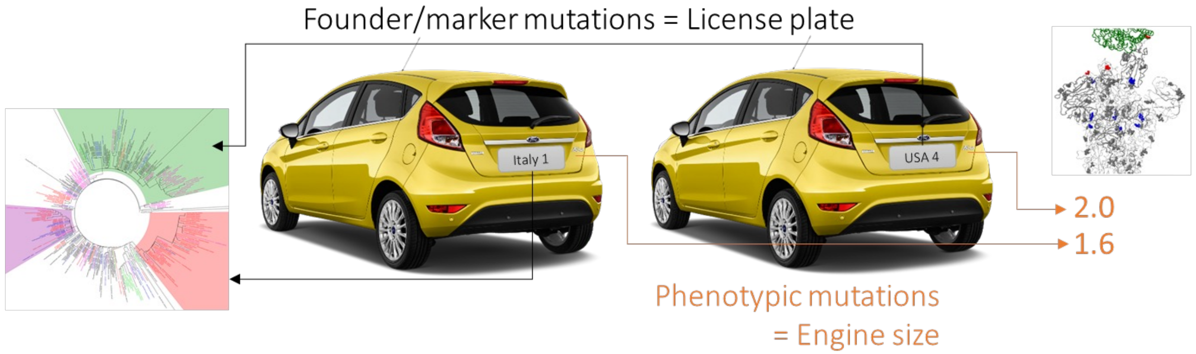
Combining relevant changes found by CoVsurver with GISAID real-time genome counts across the world allows to track and find **Emerging Variants**.

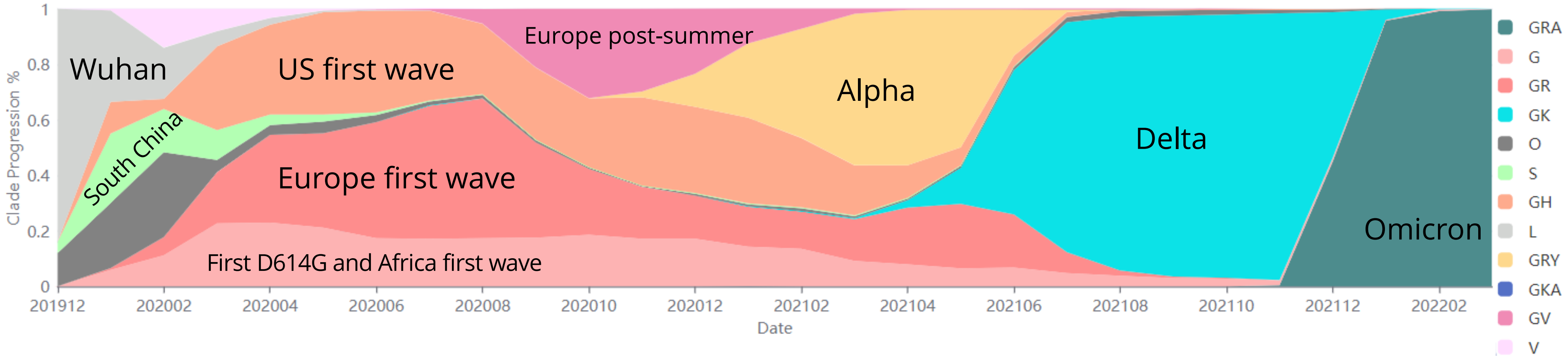
This informs CEPI, UKHSA and WHO on new trends.

POWERING DISCOVERIES

Conclusion

Our A*STAR team has worked day and night for over 2 years as pillar of GISAID and in Singapore's COVID response


Founder/marker mutations = License plate
Italy 1
USA 4
2.0
1.6
Phenotypic mutations = Engine size

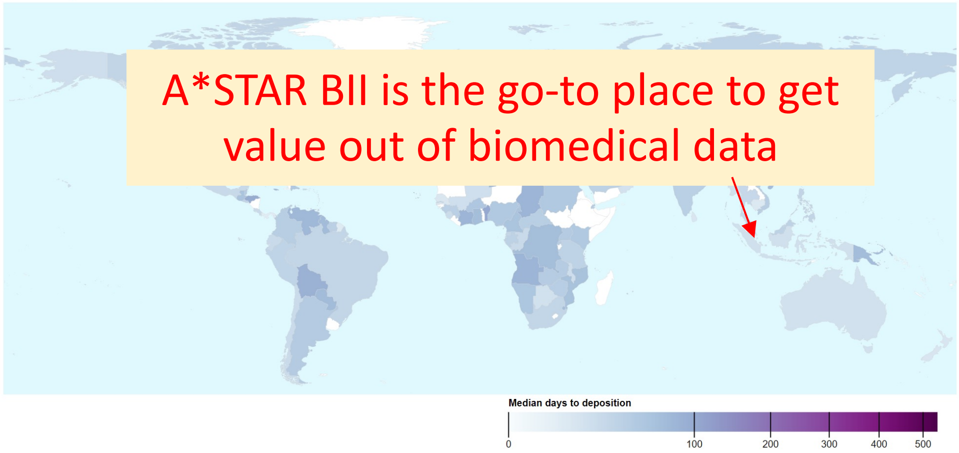


GISAID is a global data science initiative that involves public-private partnerships made possible by the GISAID community

We gratefully acknowledge all data contributors, i.e. the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative, on which this research is based.

Thank You

4,377,478 sequences with complete collection date information shared via GISAID since 28 August 2021.


Median days to deposition
0
100
200
300
400
500

A*STAR BII is the go-to place to get value out of biomedical data

POWERING DISCOVERIES