



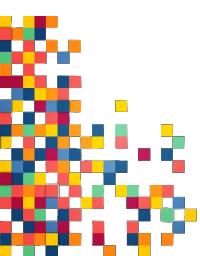
---

# Lavori su SARS-CoV-2

Gruppo Prof. Stefano Ceri

Dipartimento di Elettronica, Informazione e Bioingegneria  
Politecnico di Milano

14 Febbraio 2022



---

# Phenotype data representation

# Phenotype / clinical data representation

<http://gmql.eu/phenotype/>



The patient phenotype definition (~150 attributes) was used as a standard to collect and harmonize data from studies in the COVID-19 Host Genetics Initiative (<https://www.covid19hg.org/>)

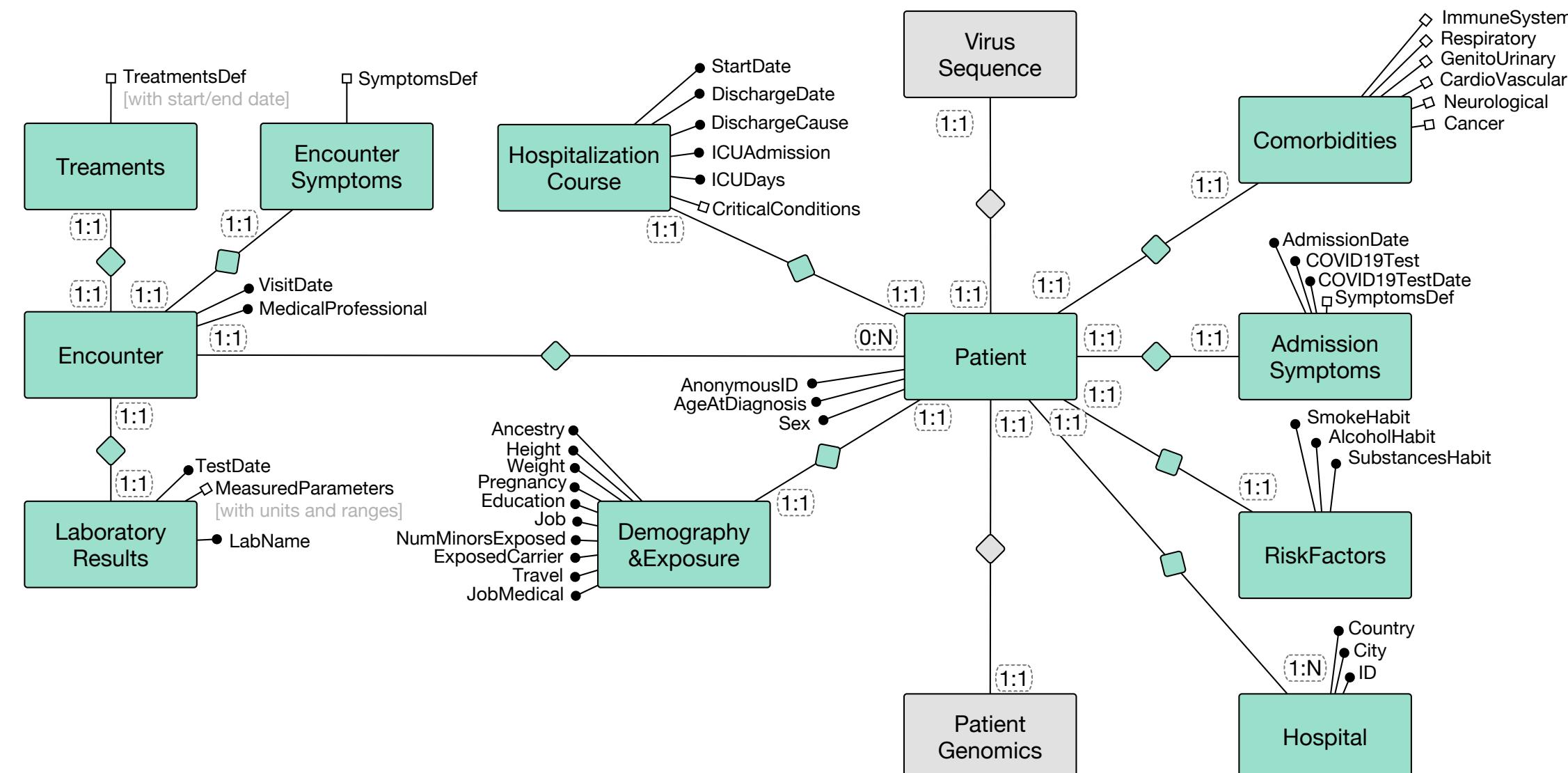
Focus on Patient and his/her data at:

- Admission
- Course of hospitalization
- Discharge

Data collected through this are hosted by the **European Genome-phenome Archive** of EMBL-EBI

COVID-19 Host Genetics Initiative Coordination,  
**Data dictionary** working group:

Stefano Ceri and **Anna Bernasconi** (Politecnico di Milano), Alessandra Renieri and Francesca Mari (Università degli Studi di Siena)

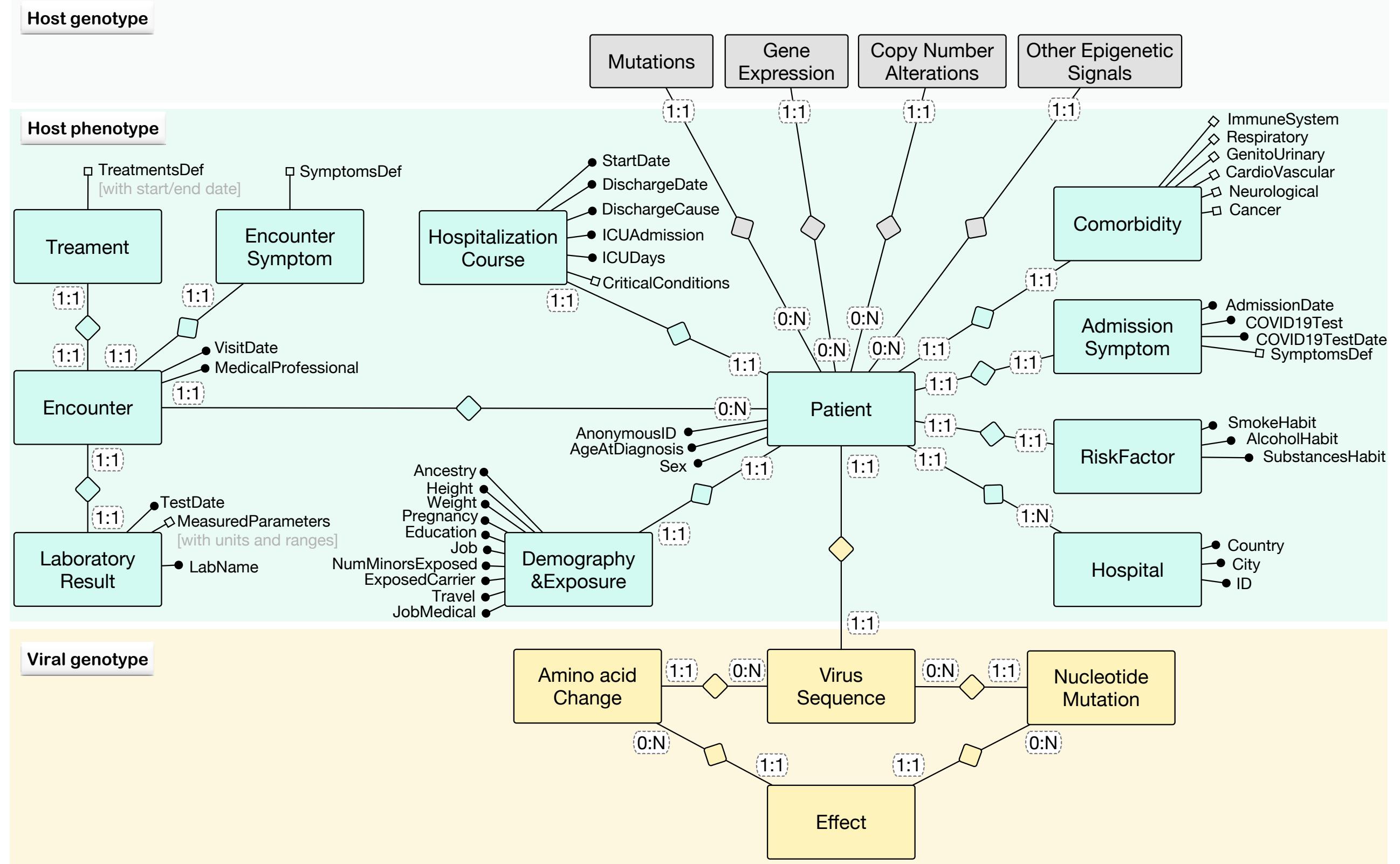


Used for studies reported in the following publications:

**COVID-19 Host Genetics Initiative (2021). Mapping the human genetic architecture of COVID-19. Nature 600, 472–477 (2021).**  
<https://doi.org/10.1038/s41586-021-03767-x>

**Van Blokland, I.V., ..., Lifelines COVID-19 cohort study, The COVID-19 Host Genetics Initiative, et al. (2021) Using symptom-based case predictions to identify host genetic factors that contribute to COVID-19 susceptibility. PloS one, 16(8), p.e0255402.** <https://doi.org/10.1371/journal.pone.0255402>

# Linking three systems



<http://gmgl.eu/genosurf/>

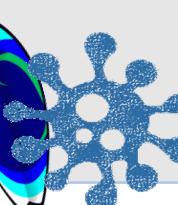


# GenoSurf

e.g., data dictionary of  
<https://www.covid19hg.org/>

# PhenotypeDB

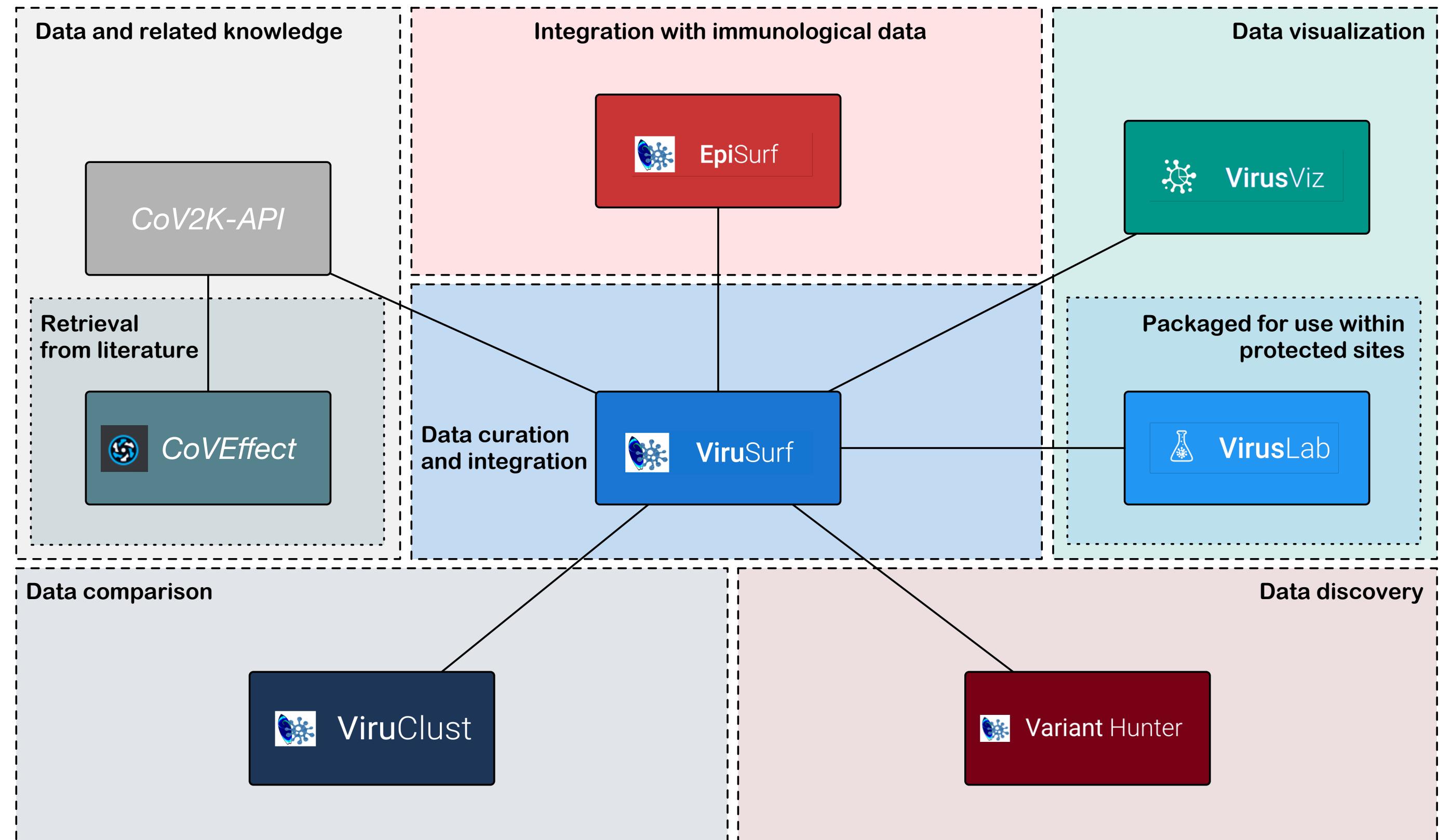
<http://gmql.eu/virusurf/>



---

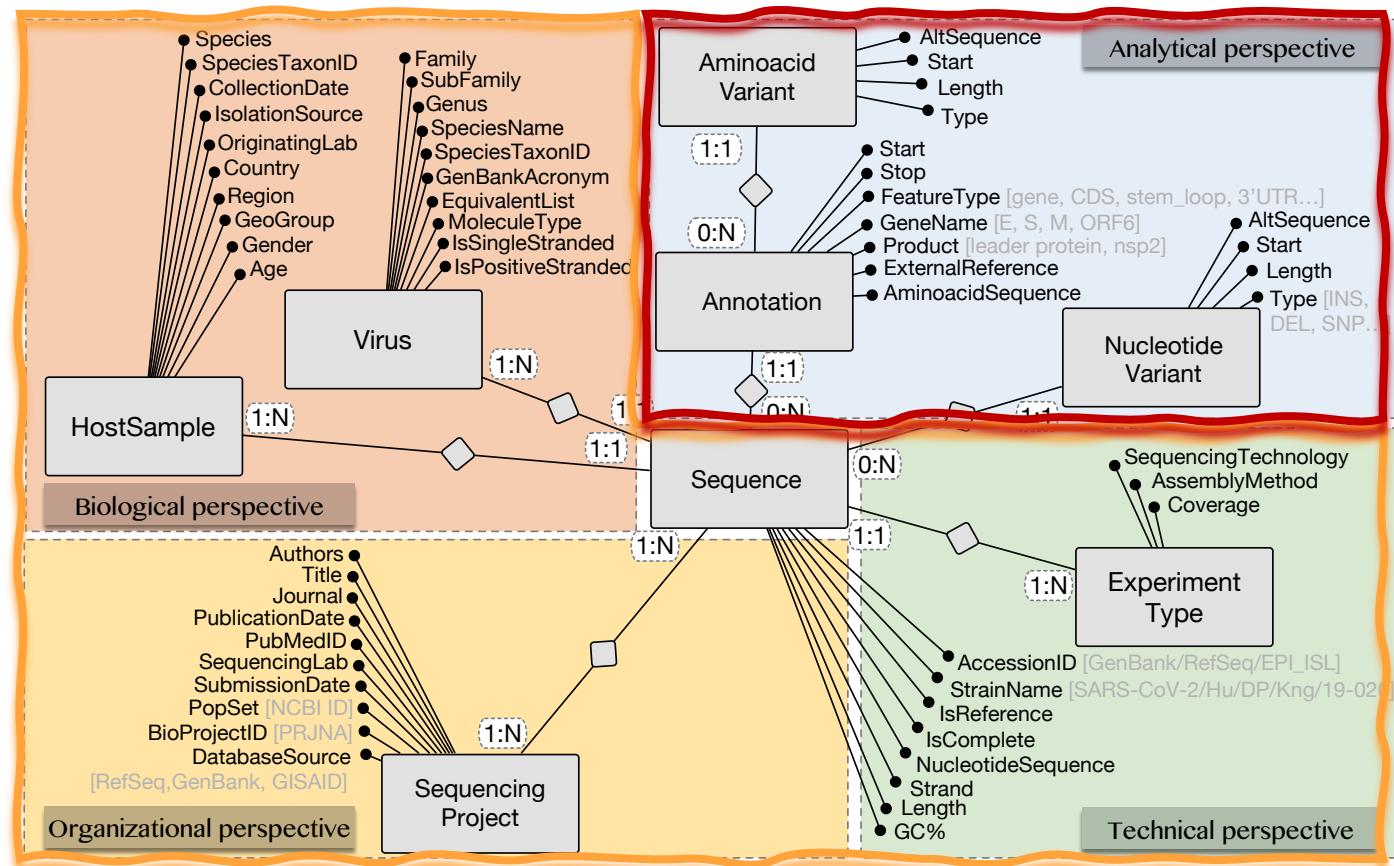
# Data Tools

# Current focus on tools for supporting SARS-CoV-2 research



# ViruSurf search system

<http://gmql.eu/virusurf/>



Source Page	Accession ID	Strain name	Is reference	Is complete	Strand	Sequence Length	GC%	N%	Lineage (Clade)	Seq. Technology	Assembly Method	Coverage	Submission date
link	England/BIRM-5F6FB/2020	England/BIRM-5F6FB/2020	False	True	positive	29903	37.99	0.42	B.1.1 (N/D)	N/D	N/D	N/D	2020-03-04
link	England/BIRM-611BF/2020	England/BIRM-611BF/2020	False	True	positive	29903	37.98	0.41	B.1.1 (N/D)	N/D	N/D	N/D	2020-03-04
link	England/BRIS-18531BS/2020	England/BRIS-18531BS/2020	False	True	positive	29903	38.0	0.4	B.1.1 (N/D)	N/D	N/D	N/D	2020-03-04

Four sections:

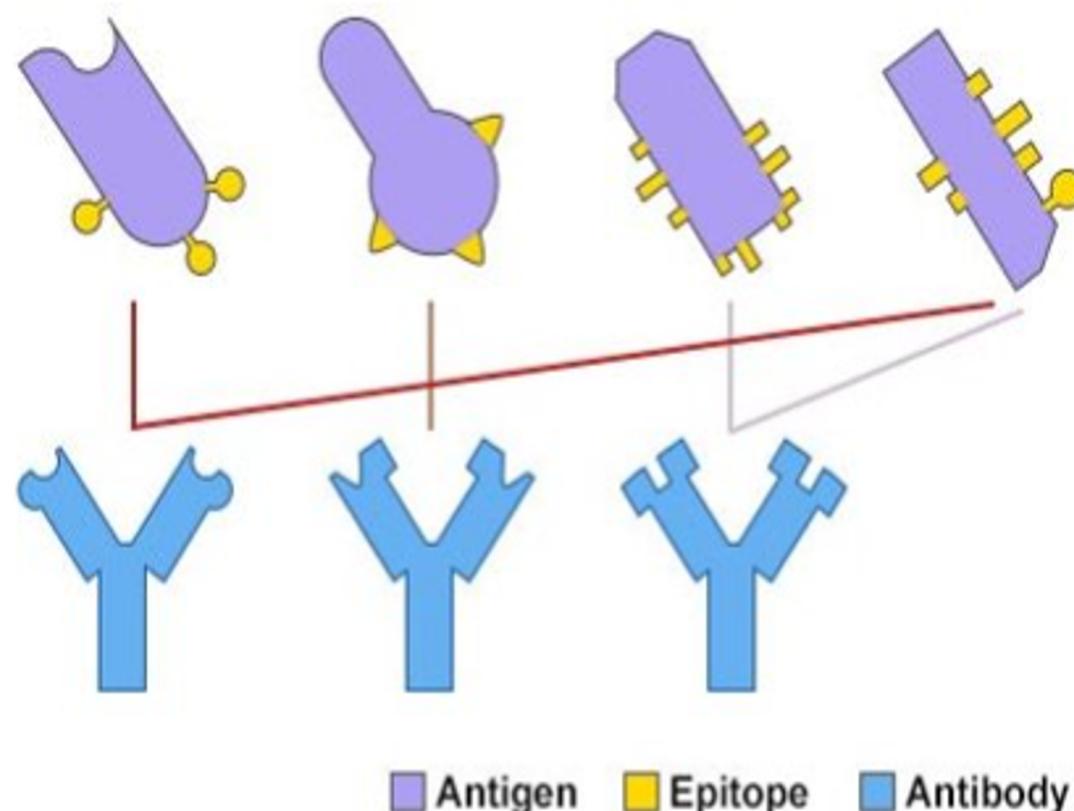
- 1) a menu bar to access the different services;
- 2) the search interface over the metadata attributes;
- 3) the search interface over annotations and nucleotide/amino acid variant information;
- 4) a result visualization section.

Canakoglu, A., Pinoli, P., Bernasconi, A., Alfonsi, T., Melidis, D. P., & Ceri, S. (2021). ViruSurf: an integrated database to investigate viral sequences. *Nucleic Acids Research*, 49(D1), D817-D824. <https://doi.org/10.1093/nar/gkaa846> - IF 16.9, SJR: Q1

# Other biological concepts: Epitopes



Epitopes are strings of amino acids from an antigen (e.g. derived from virus protein) that can be recognized by antibodies or B/T-cell receptors to provoke an immune response.



- The Immune Epitope Database (IEDB) is the largest open-source collection of epitopes for many species (over 1 million epitopes, >6K for SARS-CoV-2, ~3K refer to the Spike protein)  
Vita, R., et al., 2019. The immune epitope database (IEDB): 2018 update. Nucleic acids research, 47(D1), pp.D339-D343.
- From what is known about Pfizer and Moderna vaccines, they use all available epitopes on the Spike protein in their design.  
Jeong, D.E., et al. Assemblies of putative SARS-CoV2-spike-encoding mRNA sequences for vaccines BNT-162b2 and mRNA-1273. <https://virological.org/t/assemblies-of-putative-sars-cov2-spike-encoding-mrna-sequences-for-vaccines-bnt-162b2-and-mrna-1273/663>
- A mutation in an epitope might compromise the epitope's recognition from the immune system. E.g., E484K, E484Q, E484P are associated with the reduction of neutralization titres, possibly generating an immune escape.  
Harvey, W.T., et al. 2021. SARS-CoV-2 variants, spike mutations and immune escape. Nature Reviews Microbiology, 19(7), pp.409-424.

Source: <https://vaccsbook.com/>



## Sequence population search

The diagram illustrates the data flow into the EpiSurf system. Four external databases are shown on the left, each with a blue arrow pointing to a central server icon (a stack of three cylinders). From the server, a blue arrow points to the 'Sequence population search' interface on the right.

**Sequence population search interface:**

Novel "Severe acute respiratory syndrome coronavirus 2" sequences from "Homo sapiens" as host are preselected. If you are interested in other virus(es), please change it from the dropdown menu below:

**Sequence population search condition:** taxon\_name: ["severe acute respiratory syndrome coronavirus 2"], host\_taxon\_name: ["homo sapiens"]

Virus	
Virus taxon ID	Virus taxon name severe acute respiratory syndrome coronavirus 2
Virus species	

Host Organism					
Host taxon name homo sapiens	Collection date	Isolation source	Continent	Country	Region
Gender	Age				

Sequence properties and technology					
Is reference	Is complete	Strand	Sequence Len...	GC%	N%
Lineage	Sequencing techn...	Assembly method	Coverage		

Organization			
Submitting Lab	Submission date	BioProject ID	Database source

## Epitope search

Custom Epitopes    USE IEDB EPITOPE WITHOUT VARIANT COUNTS    USE IEDB EPITOPE WITH VARIANT COUNTS

**IEDB Epitope search**

**Epitope search condition:** protein: ["Spike (surface glycoprotein)"], assay\_type: ["B cell"]

Protein Name Spike (surface glycoprotei	Assay B cell	HLA restriction	Is Linear true
			724
Response Frequency	Position Range	Epitope IEDB ID	false
			290

**CLEAR EPITOPE QUERY**

**Bernasconi, A., Ciliberti, L., Al Khalaf, R., Alfonsi, T., Ceri, S., Pinoli, P., & Canakoglu, A. (2021) EpiSurf: metadata-driven search server for analyzing amino acid changes on epitopes of SARS-CoV-2 and other viral species. Database, 2021. <https://doi.org/10.1093/database/baab059> – IF: 3.5, SJR: Q1**

# Checking the Delta mutations in UK over important epitope ranges

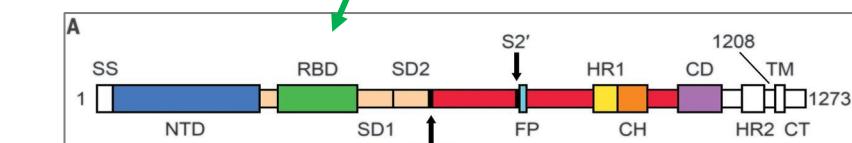


## Spike RBD mutations and immune escape

151,559 sequences

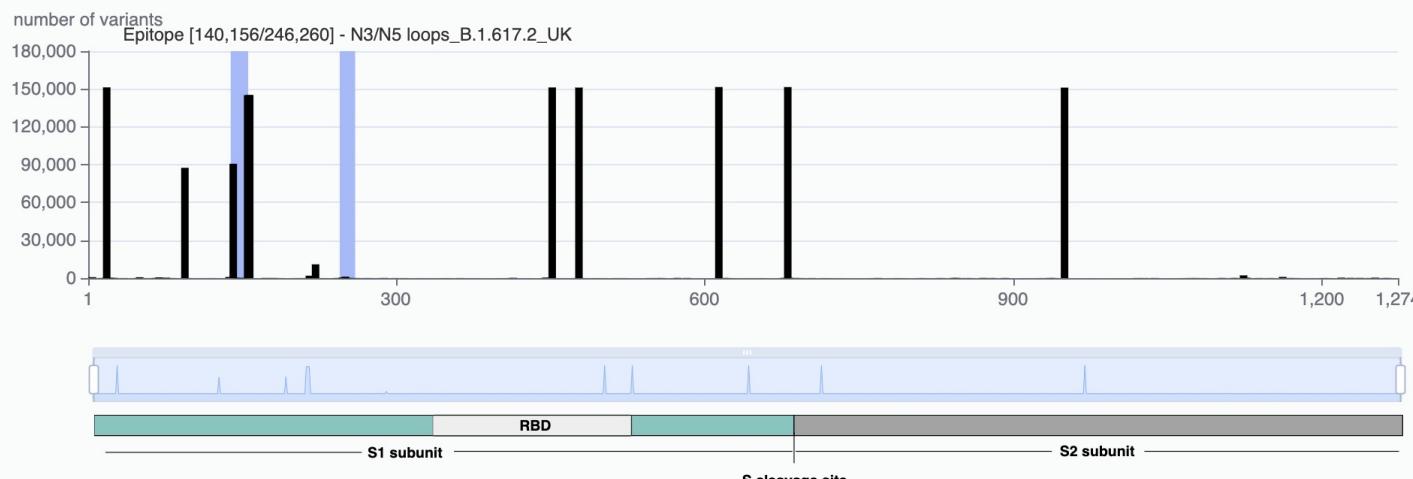


Region<sup>i</sup> 319 – 541 Receptor-binding domain (RBD)

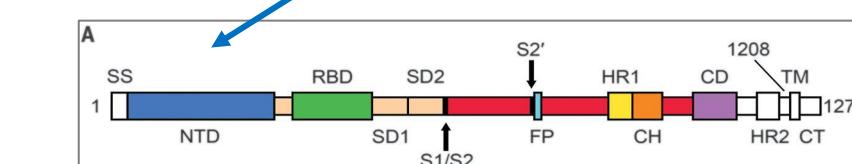


## Spike NTD mutations and immune escape

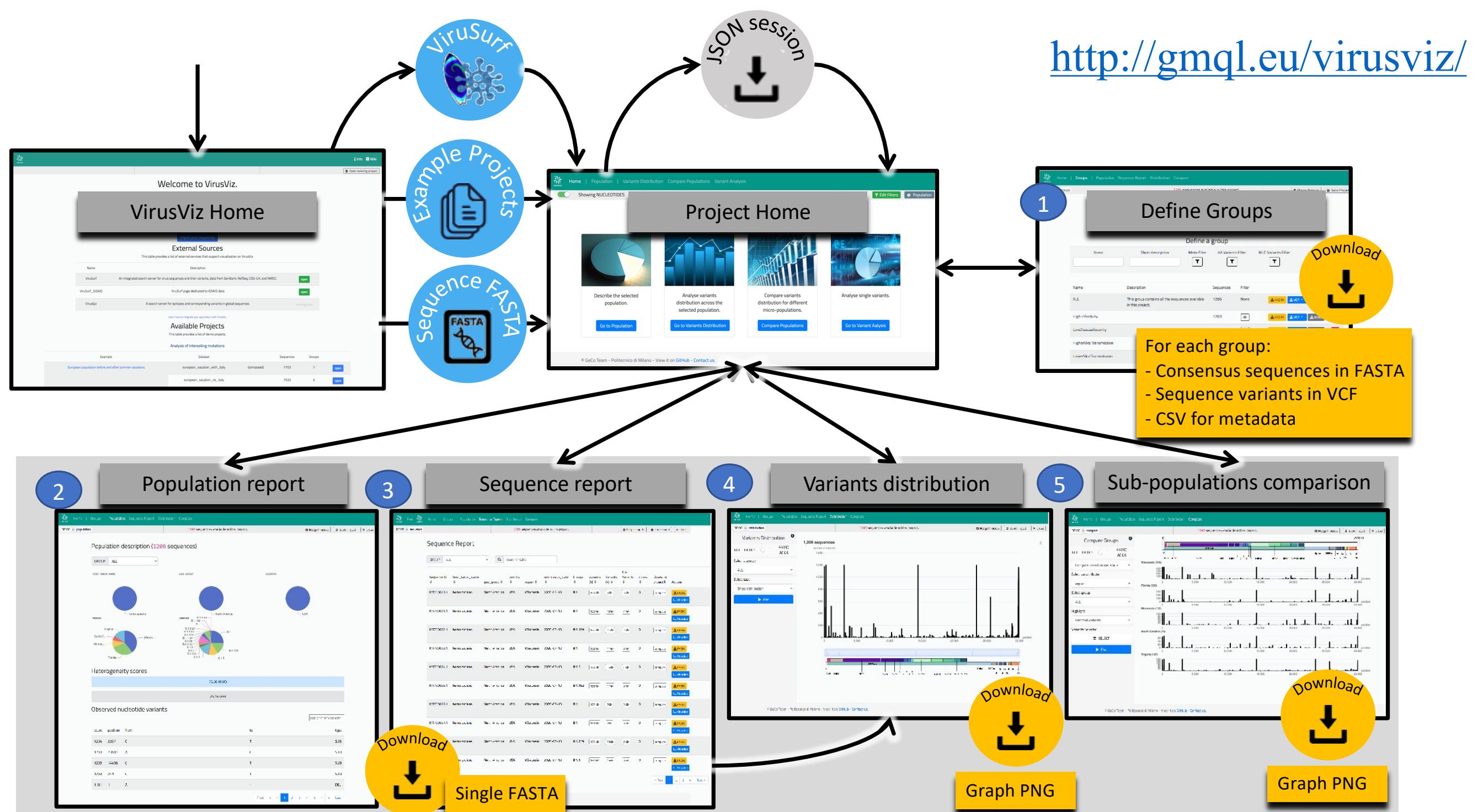
151,559 sequences



Domain<sup>i</sup> 13 – 303 BetaCoV S1-NTD

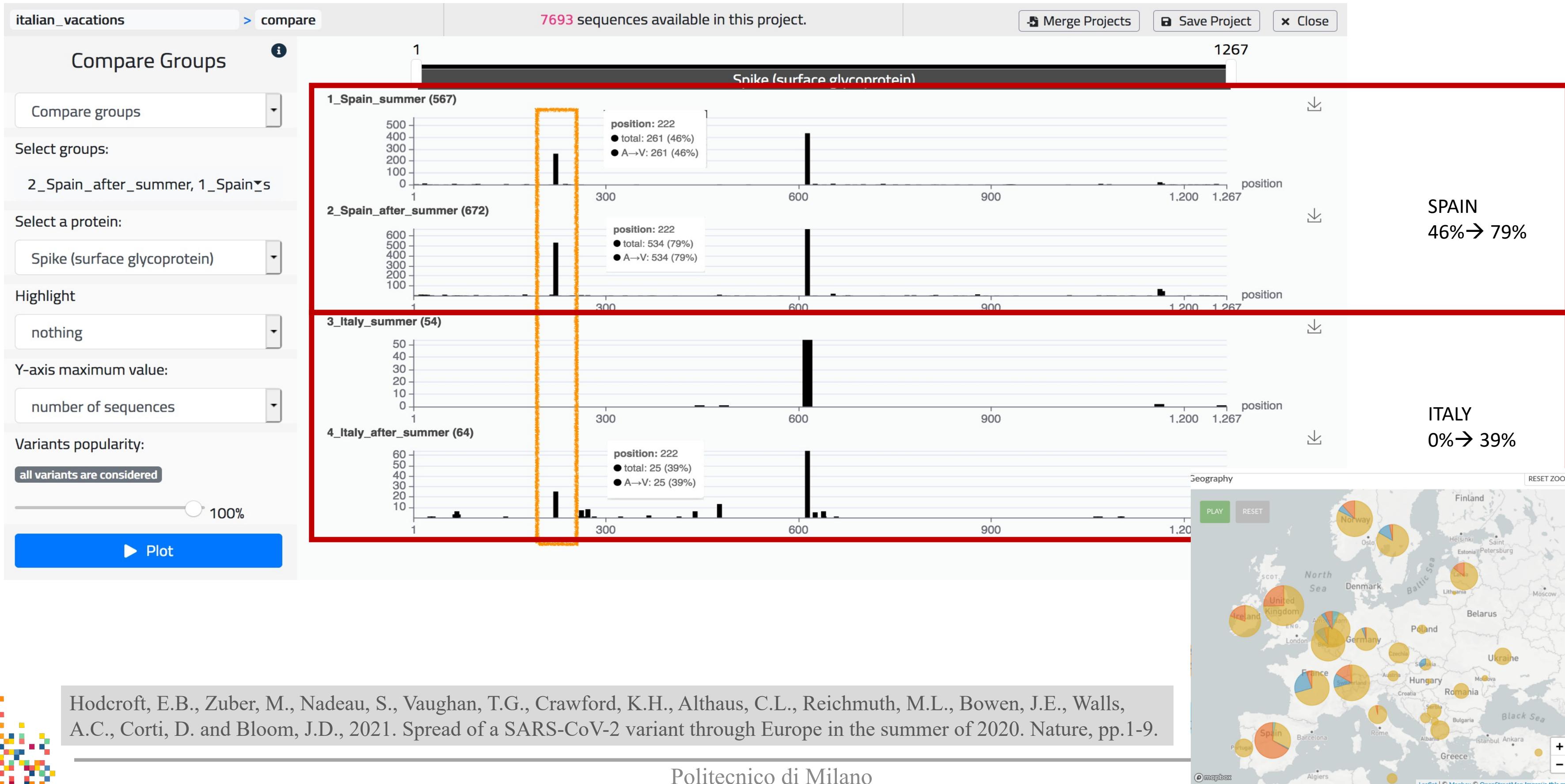


# VirusViz: comparative visualization of nucleotide/amino acid mutations



Bernasconi, A., Gulino, A., Alfonsi, T., Canakoglu, A., Pinoli, P., Sandionigi, A., & Ceri, S. (2021) VirusViz: Comparative analysis and effective visualization of viral nucleotide and amino acid variants. Nucleic Acids Research, 49(15), e90. <https://doi.org/10.1093/nar/gkab478> - IF 16.9, SJR: Q1

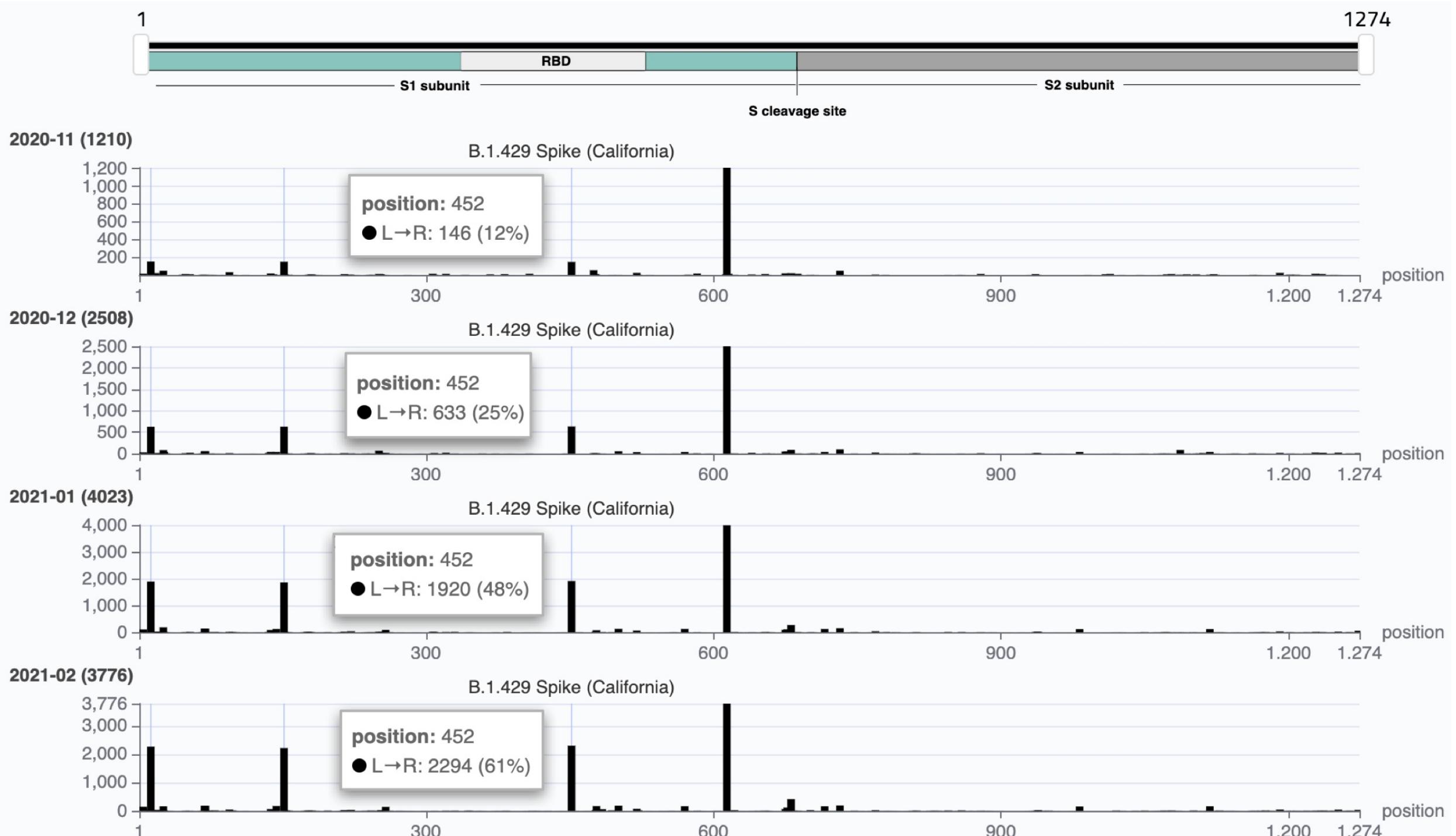
# The Spike mutation A222V arrived from Spain after Summer 2020



# Californian variant (Epsilon)



The first alert of the “Epsilon variant” was in a letter dated Feb. 11, 2021, indicating 3 amino acid changes on Spike: S13I, W152C, L452R.



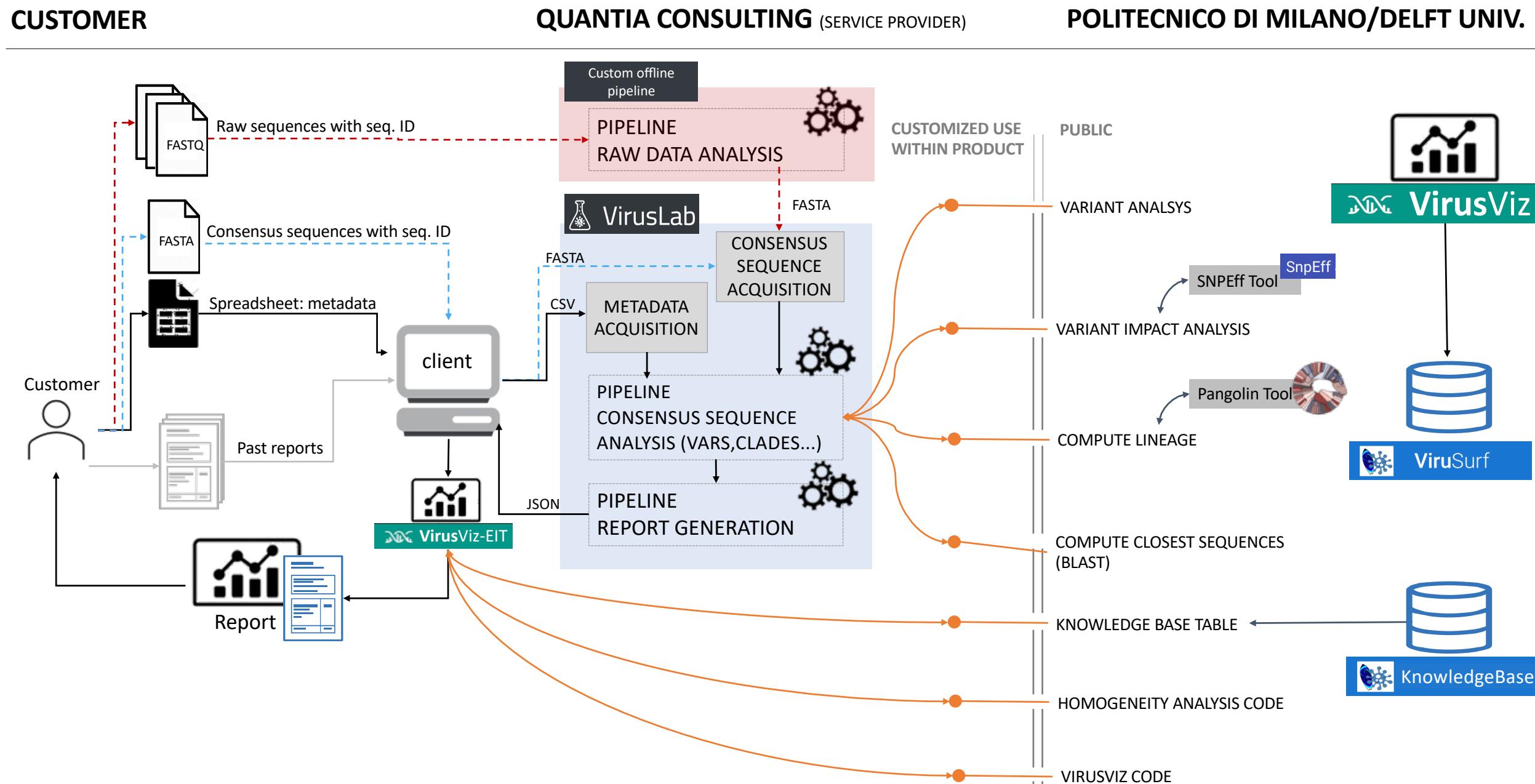
Zhang, W., Davis, B.D., Chen, S.S., Martinez, J.M.S., Plummer, J.T. and Vail, E., 2021. Emergence of a novel SARS-CoV-2 variant in Southern California. *Jama*, 325(13), pp.1324-1326.

# VirusLab: Packaging of customizable services for use within protected sites



For supporting a laboratory wishing to perform secondary (raw data) analysis and to add sensible metadata (e.g. clinical), still using our data and knowledge bases and visualization tools

<http://viruslab.quantiaconsulting.com/viruslab/>



Project supported by:  
EIT "DATA against COVID-19"  
Innovation Activity, Project  
20663 "ViruSurf"



Pinoli, P., Bernasconi, A., Sandionigi, A., & Ceri, S. VirusLab: a tool for customized SARS-CoV-2 data analysis. BioTech 2021, 10(4), 27.  
<https://doi.org/10.3390/biotech10040027>



Supporting pairwise comparison (target/background) of viral populations in

- **lineages** (with different mutations)
- **time** (time intervals vs monthly/weekly periods)
- **space** (different continents, countries, regions)
- **any of the above** (custom analyses)

Integration of different types of analyses allows the generation of “testable hypotheses” that can be used to pinpoint interesting evolutionary patterns

The screenshot shows the ViruClust web application interface. At the top, there is a dark header bar with the ViruClust logo, the text "enabled by data from **GISAID**", and a "Last update date: 2021-11-22" timestamp. Below the header is a navigation bar with five tabs: "PREVALENCE OF LINEAGES" (selected), "EVOLUTION IN TIME", "EVOLUTION IN SPACE", "CUSTOM ANALYSIS", and a collapsed menu icon. The main content area is titled "HOW TO USE VIRUCLUST". It contains four sections, each with a title and a "START THIS ANALYSIS" button:

- PREVALENCE OF LINEAGES
- EVOLUTION IN TIME
- EVOLUTION IN SPACE
- CUSTOM ANALYSIS

Cilibriasi, L., Pinoli, P., Bernasconi, A., Canakoglu, A., Chiara, M., & Ceri, S. ViruClust: direct comparison of SARS-CoV-2 genomes and genetic variants in space and time. Accepted by the Bioinformatics Journal – IF: 6.9, SJR: Q1 [*in print*]

# ViruClust application: Early evolution of Delta



Are there mutations not observed in India (place of origin) that are instead observed in the rest of Asia?

## Spatial analysis [India (target) vs. rest of Asia (background)]

- Results highlight 5 mutations observed in the spike protein of the Delta variant in India but with a much higher frequency outside of India

## Temporal analysis [3 months (May to July 2021)]

- In India: the 5 changes never reach >90% prevalence, with different, increasing, profiles of frequency over time
- In the rest of Asia: the 5 changes are fixed at >90% prevalence; only one G142D show a detectable increase

If the “novel” mutations are “selected” and confer some advantages to Delta, they should not be observed in other, closely related, variants (e.g., Kappa)

## Custom analysis [Delta vs Kappa]

- Delta shows Spike:E156G, F157-, R158-, and D950N
  - Kappa did not have these (only G142D)
- ... suggesting that the acquisition of novel/improved mutations can have epidemiological implications

TABLE ④

mutation	p_value ↓ ③	odds_ratio	%_target ↓ ②	%_background ↓ ①
Spike_D614G	0.13824	0.99970	99.87255 % (15673)	99.90291 % (9261)
Spike_P681R	0.00000	0.98656	98.40056 % (15442)	99.74110 % (9246)
Spike_T478K	0.00000	0.95328	94.52622 % (14834)	99.15858 % (9192)
Spike_L452R	0.00000	0.95888	95.02963 % (14913)	99.10464 % (9187)
Spike_T19R	0.00000	0.94584	93.56401 % (14683)	98.92125 % (9170)
Spike_D950N	0.00000	0.77111	73.24285 % (11494)	94.98382 % (8805)
Spike_E156G	0.00000	0.33527	30.88638 % (4847)	92.12513 % (8540)
Spike_F157-	0.00000	0.33648	30.99471 % (4864)	92.11435 % (8539)
Spike_R158-	0.00000	0.33797	31.10304 % (4881)	92.02805 % (8531)
Spike_G142D	0.00000	0.33431	25.67387 % (4029)	76.79612 % (7119)





Supporting two working modes:

- Lineage **independent** mutation growth
- Lineage **dependent** mutation growth

The screenshot shows the Variant Hunter web application interface. At the top, there is a dark red header bar with the title "Variant Hunter" and a logo on the left. To the right of the title is a menu icon consisting of three horizontal lines. Below the header, there are two red buttons: "LINEAGE INDEPENDENT" on the left and "LINEAGE SPECIFIC" on the right. The main content area has a yellow background. In the center, there is a red rectangular form titled "DEFINE ANALYSIS:". This form contains several input fields:

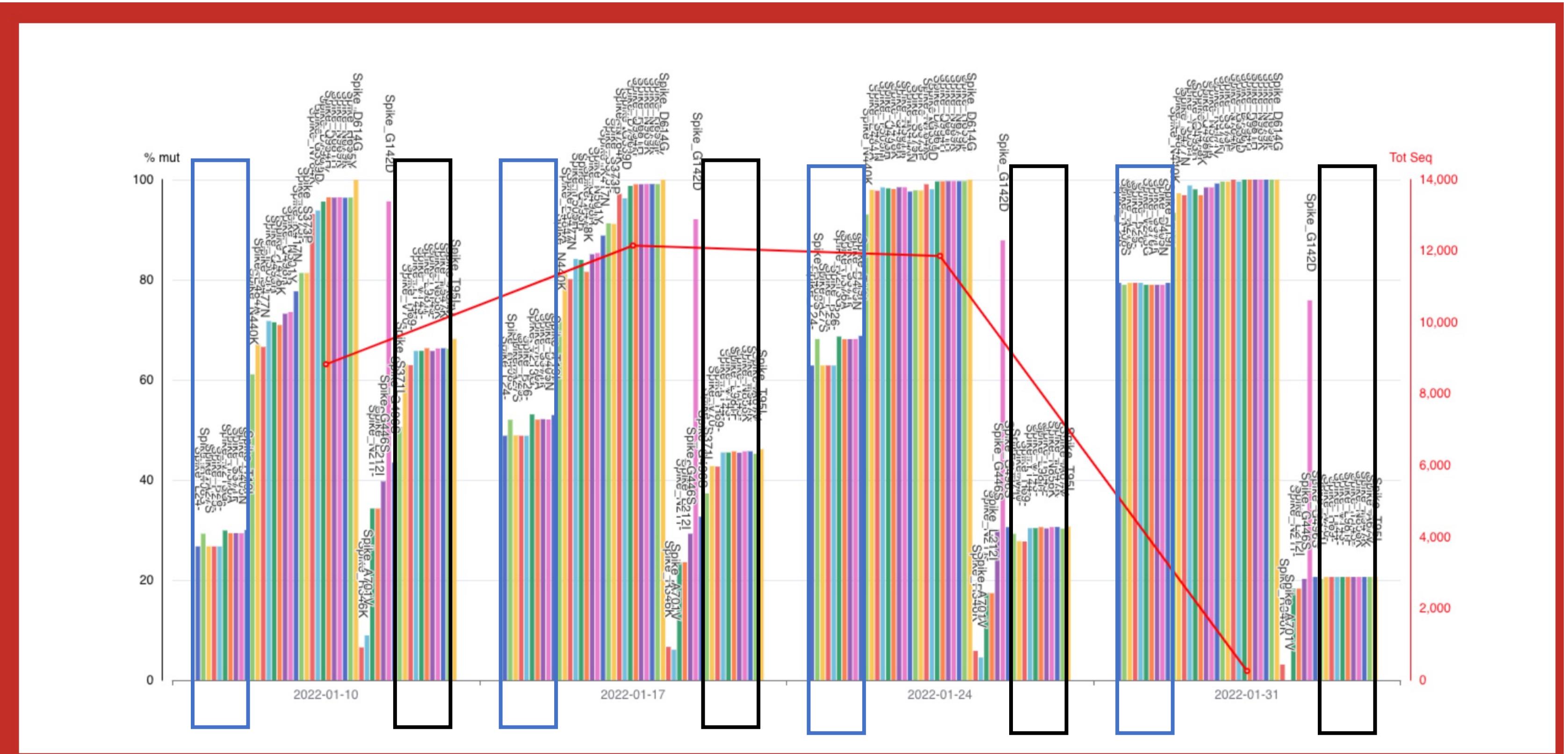
- "Granularity:" dropdown set to "world".
- "Location:" dropdown set to "Place".
- "Date:" input field showing "2022-01-08" with a calendar icon.
- "# Week:" dropdown set to "4".
- A "START ANALYSIS" button at the bottom.

# Mutations that are growing fast in Denmark

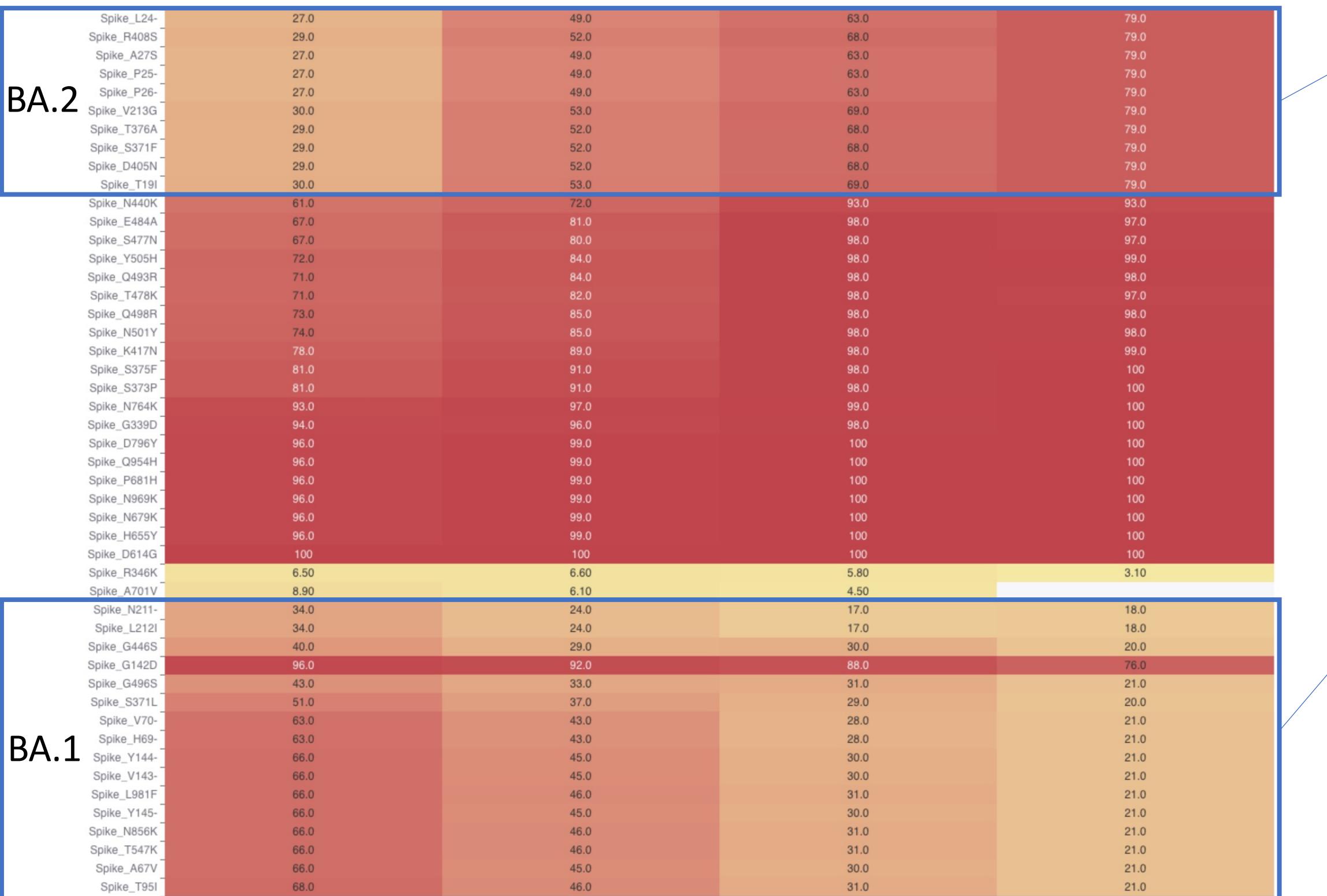


Location	Protein	Mut	Slope ↓ 1	Y-intercept	Presence (%) 2022-01-10	Presence (%) 2022-01-17	Presence (%) 2022-01-24	Presence (%) 2022-01-31	P-value with mut	P-value without mut
Denmark	Spike	L24-	17	28.64	26.71 (2360)	48.83 (5936)	62.86 (7460)	79.38 (204)	0.0	0.0
Denmark	Spike	R408S	17	32.30	29.26 (2585)	52.02 (6323)	68.15 (8087)	78.99 (203)	0.0	0.0
Denmark	Spike	A27S	17	28.69	26.73 (2362)	48.91 (5945)	62.88 (7462)	79.38 (204)	0.0	0.0
Denmark	Spike	P25-	17	28.64	26.71 (2360)	48.83 (5936)	62.86 (7460)	79.38 (204)	0.0	0.0
Denmark	Spike	P26-	17	28.64	26.71 (2360)	48.83 (5936)	62.86 (7460)	79.38 (204)	0.0	0.0
Denmark	Spike	V213G	16	33.24	29.90 (2642)	53.09 (6454)	68.65 (8147)	78.99 (203)	0.0	0.0
Denmark	Spike	T376A	16	32.38	29.36 (2594)	52.04 (6326)	68.11 (8083)	78.99 (203)	0.0	0.0
Denmark	Spike	S371F	16	32.43	29.37 (2595)	52.14 (6338)	68.13 (8085)	78.99 (203)	0.0	0.0
Denmark	Spike	D405N	16	32.39	29.35 (2593)	52.08 (6331)	68.16 (8088)	78.99 (203)	0.0	0.0
Denmark	Spike	T19I	16	33.19	30.01 (2651)	52.96 (6438)	68.79 (8163)	79.38 (204)	0.0	0.0
Denmark	Spike	N440K	12	62.18	61.10 (5398)	71.96 (8748)	93.03 (11040)	93.39 (240)	3.5e-201	0.0
Denmark	Spike	E484A	11	69.89	67.16 (5934)	81.33 (9886)	97.95 (11624)	97.28 (250)	8.3e-203	0.0

# BA.1 vs BA.2 (Denmark)



# BA.1 vs BA.2 (Denmark)



PROTEIN	BA1	BA.2
<b>SPIKE</b>		S:T19I
		S:L24-
		S:P25-
		S:P26-
		S:A27S
	S:A67V	
	S:H69-	
	S:V70-	
	S:T95I	
	S:G142-	S:G142D
	S:V143-	
	S:Y144-	
	S:Y145D	
	S:N211-	
	S:L212I	
		S:V213G
		S:G339D
		S:S371F
		S:S373P
		S:S375F
		S:T376A
		S:D405N
		S:R408S
	S:K417N	S:K417N
	S:N440K	S:N440K
	S:G446S	
	S:S477N	S:S477N
	S:T478K	S:T478K
	S:E484A	S:E484A
	S:Q493R	S:Q493R
	S:G496S	
	S:Q498R	S:Q498R
	S:N501Y	S:N501Y
	S:Y505H	S:Y505H
	S:T547K	
	S:D614G	S:D614G
	S:H655Y	S:H655Y
	S:N679K	S:N679K
	S:P681H	S:P681H
	S:N764K	S:N764K
	S:D796Y	S:D796Y
	S:N856K	
	S:Q954H	S:Q954H
	S:N969K	S:N969K
	S:L981F	



# Growing mutations of BA.1 in United Kingdom at the end of 2021



Sorted by descending slope

United Kingdom / BA.1 / 2021-12-31 / 4 weeks

TABLE ⓘ

-	Location	Protein	Mut	Slope ↓ 1	Y-intercept	Presence (%) 2021-12-10	Presence (%) 2021-12-17	Presence (%) 2021-12-24	Presence (%) 2021-12-31
□	United Kingdom	Spike	G339D	29	12.50	7.82 (4976) (27843)	45.12 (27843)	75.58 (30195)	92.88 (13352)
□	United Kingdom	Spike	N969K	29	12.54	7.82 (4980) (27881)	45.19 (27881)	75.70 (30241)	92.89 (13354)
□	United Kingdom	N	R32-	29	12.42	7.78 (4953) (27736)	44.95 (27736)	75.67 (30230)	92.88 (13353)
□	United Kingdom	Spike	G496S	29	11.20	7.18 (4570) (26744)	43.34 (26744)	74.07 (29593)	92.84 (13347)
□	United Kingdom	Spike	Q954H	29	12.54	7.82 (4980) (27885)	45.19 (27885)	75.70 (30241)	92.90 (13355)
□	United Kingdom	Spike	Y144-	29	12.30	7.81 (4973) (27604)	44.74 (27604)	74.92 (29930)	92.80 (13341)
□	United Kingdom	Spike	Q493R	29	11.24	7.20 (4582) (26790)	43.42 (26790)	74.08 (29596)	92.89 (13354)
□	United Kingdom	NSP6	I189V	29	12.61	7.88 (5016) (27911)	45.23 (27911)	75.70 (30244)	92.85 (13348)
□	United Kingdom	E	T9I	29	12.34	7.77 (4946) (27651)	44.81 (27651)	75.51 (30166)	92.89 (13354)
□	United Kingdom	NSP6	G107-	29	12.52	7.80 (4965) (27871)	45.17 (27871)	75.66 (30228)	92.88 (13352)
□	United Kingdom	Spike	N679K	29	12.54	7.82 (4980) (27878)	45.18 (27878)	75.68 (30235)	92.86 (13350)
□	United Kingdom	NSP3	K38R	29	12.52	7.83 (4984) (27839)	45.12 (27839)	75.62 (30211)	92.85 (13348)

# Growing mutations in United Kingdom compared to Omicron (BA.1)



United Kingdom / BA.1 / 2021-12-31 / 4 weeks ^

**TABLE** 

Location	Lineage	Protein	Mut	Slope	Y-intercept	P-value with mut	P-value without mut	P-value comparative ↑ 1
United Kingdom	BA.1	Spike	N764K	-1.9	55	0.0	0.0	0.0
United Kingdom	BA.1	Spike	R346K	2.6	6.3	0.0	0.0	1.2e-92
United Kingdom	BA.1	NS3	L106F	-1.7	18	0.0	0.0	2.4e-24
United Kingdom	BA.1	N	D343G	-1.6	18	0.0	0.0	5.3e-23
United Kingdom	BA.1	Spike	I1081V	-0.47	11	0.0	0.0	5.2e-12
United Kingdom	BA.1	NSP3	S1265N	-0.89	3.9	0.0	0.0	1.4e-7
United Kingdom	BA.1	NSP3	Y129H	0.32	2.0	0.0	0.0	0.0000011
United Kingdom	BA.1	NSP12	D153Y	0.29	1.6	0.0	0.0	0.0000028
United Kingdom	BA.1	Spike	A701V	-0.67	27	0.0	0.0	0.000019
United Kingdom	BA.1	NSP3	V1069I	-0.66	34	0.0	0.0	0.000043

**P value with mut:** shows if the population «lineage + mutation» is growing differently compared to everything.

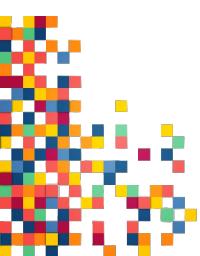
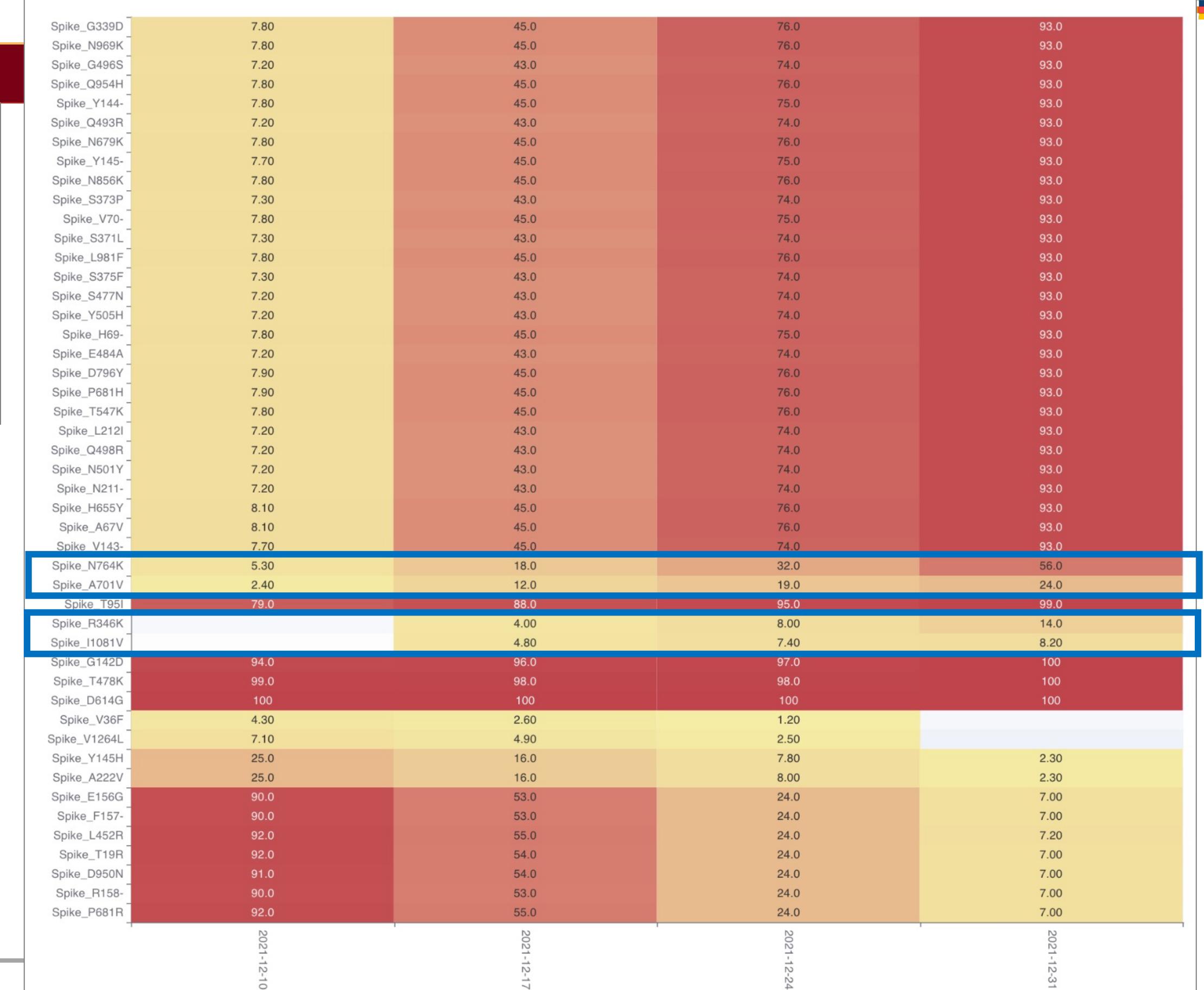
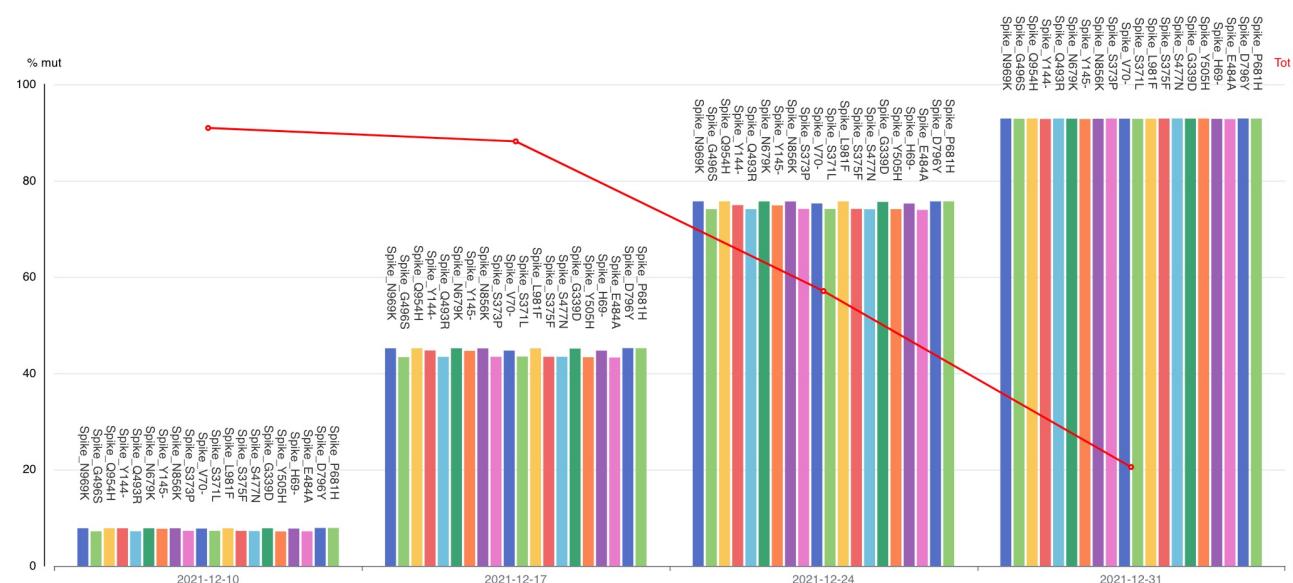
**P value without mut:** shows if the population «lineage without mutation» is growing differently compared to everything.

**P value comparative:** shows if the population «lineage + mutation» is growing differently compared to the population «lineage without mutation».

# Relevant Spike mutations in United Kingdom at the end of 2021



United Kingdom / BA.1 / 2021-12-31 / 4 weeks



# Focus on: Spike mutations in United Kingdom compared to Omicron (BA.1)



United Kingdom / BA.1 / 2021-12-31 / 4 weeks

**TABLE**

Location	Lineage	Protein	Mut	Slope	Y-intercept	P-value with mut	P-value without mut	P-value comparative ↑
United Kingdom	BA.1	Spike	N764K	-1.9	55	0.0	0.0	0.0
United Kingdom	BA.1	Spike	R346K	2.6	6.3	0.0	0.0	1.2e-92
United Kingdom	BA.1	Spike	I1081V	-0.47	11	0.0	0.0	5.2e-12
United Kingdom	BA.1	Spike	A701V	-0.67	27	0.0	0.0	0.000019

Rows per page: 10 ▾ 1-4 of 4 < >

The observations of the third p-value and of the heatmap indicate that the Omicron lineage with the additional Spike mutation R346K is growing significantly faster than Omicron alone

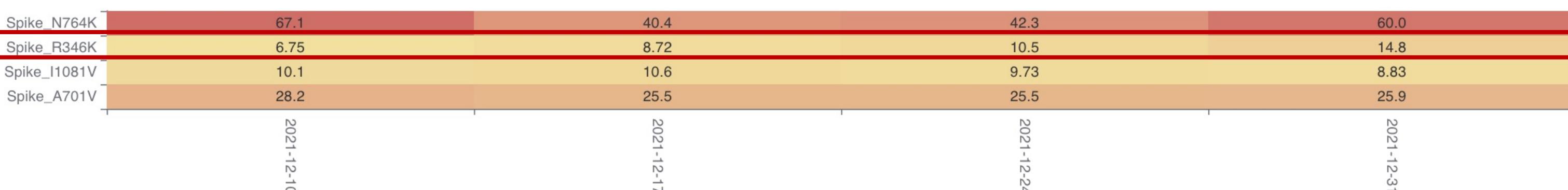
Spike\_R346K mutation is currently under scrutiny...  
... but so far its addition to Omicron did not affect the neutralization susceptibility

> Clin Infect Dis. 2021 Dec 16;ciab1041. doi: 10.1093/cid/ciab1041. Online ahead of print.

## Neutralization of SARS-CoV-2 Omicron variant by sera from BNT162b2 or Coronavac vaccine recipients

Lu Lu <sup>1</sup>, Bobo Wing-Yee Mok <sup>1</sup>, Lin-Lei Chen <sup>1</sup>, Jacky Man-Chun Chan <sup>2</sup>, Owen Tak-Yin Tsang <sup>2</sup>, Bosco Hoi-Shiu Lam <sup>3</sup>, Vivien Wai-Man Chuang <sup>4</sup>, Allen Wing-Ho Chu <sup>1</sup>, Wan-Mui Chan <sup>1</sup>, Jonathan Daniel Ip <sup>1</sup>, Brian Pui-Chun Chan <sup>1</sup>, Ruiqi Zhang <sup>5</sup>, Cyril Chik-Yan Yip <sup>1</sup> <sup>6</sup>, Vincent Chi-Chung Cheng <sup>1</sup> <sup>6</sup>, Kwok-Hung Chan <sup>1</sup>, Dong-Yan Jin <sup>7</sup>, Ivan Fan-Ngai Hung <sup>5</sup>, Kwok-Yung Yuen <sup>1</sup> <sup>6</sup>, Honglin Chen <sup>1</sup>, Kelvin Kai-Wang To <sup>1</sup> <sup>6</sup>

**P value with mut:** shows if the population «lineage + mutation» is growing differently compared to everything else  
**P value without mut:** shows if the population «lineage without mutation» is growing differently compared to everything else  
**P value comparative:** shows if the population «lineage + mutation» is growing differently compared to the population «lineage without mutation»



# The IHU variant from Cameroon



**AFRICA  
TIMES**

POLITICS & POLICY • LEADERS & COMPANIES • FINANCE & ECON  
SECURITY • HEALTH • CUTTING EDGE • YOUNG VOICES • AFR

News Articles Commentary Blogs In Depth

HOME > HEALTH > FRENCH SCIENTISTS REPORT NEW COVID VARIANT TRACED TO CAMEROON

**French scientists report new COVID variant traced to Cameroon**

By AT editor - 3 January 2022 at 9:07 pm

🕒 06 Jan

**Covid-19: France detects new variant called IHU in man who travelled to Cameroon**

**news24** Lenin Ndebele

SHARE

**Comments (5)**

**Emergence in Southern France of a new SARS-CoV-2 variant of probably Cameroonian origin harbouring both substitutions N501Y and E484K in the spike protein**

Philippe Colson, Jérémie Delerue, Emilie Burel, Jordan Dahan, Agnès Jouffret, Florence Fenollar, Nouara Yahi, Jacques Fantini, Bernard La Scola, Didier Raoult

doi: <https://doi.org/10.1101/2021.12.24.21268174>

This article is a preprint and has not been peer-reviewed [what does this mean?]. It reports new medical research that has yet to be evaluated and so should not be used to guide clinical practice.

**Abstract** Full Text Info/History Metrics

**ABSTRACT**

SARS-CoV-2 variants have become a major virological, epidemiological and clinical concern, particularly with regard to the risk of escape from vaccine-induced immunity. Here we describe the emergence of a new variant. For twelve SARS-CoV-positive patients living in the same geographical area of southeastern France, qPCR testing that screen for variant-associated mutations showed an atypical combination. The index case returned from a travel in Cameroon. The genomes were obtained by next-generation sequencing with Oxford Nanopore Technologies on GridION instruments within ≈8 h. Their analysis revealed 46 mutations and 37 deletions resulting in 30 amino acid substitutions and 12 deletions. Fourteen amino acid substitutions, including N501Y and E484K, and 9 deletions are located in the spike protein. This genotype pattern led to create a new Pangolin lineage named B.1.640.2, which is a phylogenetic sister group to the old B.1.640 lineage renamed B.1.640.1. Both lineages differ by

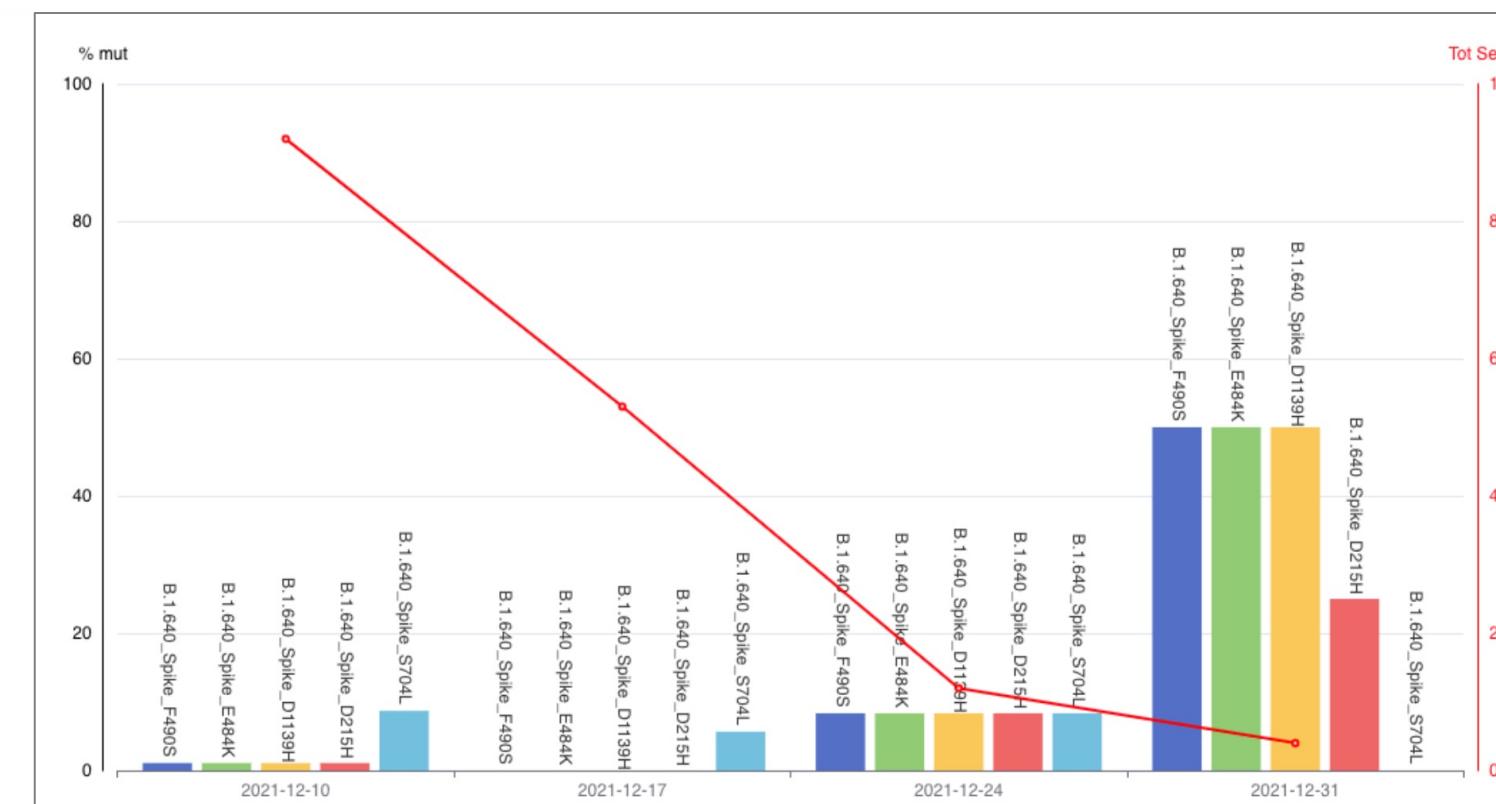
# B.1.640 mutations «run faster» and create the B.1.640.2 lineage



/ B.1.640 / 2021-12-31 / 4 weeks

TABLE ⓘ

Location	Lineage	Protein	Mut	Slope	Y-intercept	P-value with mut	P-value without mut	P-value comparative ↑ 1
World	B.1.640	Spike	F490S	15	-8.0	0.064	3.6e-9	0.000072
World	B.1.640	Spike	E484K	15	-8.0	0.064	3.6e-9	0.000072
World	B.1.640	Spike	D1139H	15	-8.0	0.064	3.6e-9	0.000072
World	B.1.640	Spike	D215H	7.9	-3.0	0.41	7.8e-9	0.047
World	B.1.640	Spike	S704L	-2.0	9.0	0.27	9.2e-7	0.93



## B.1.640.1 and B.1.640.2 lineages

B.1.640.1	B.1.640.2 (IHU variant)
S:P9L	S:P9L
S:E96Q	S:E96Q
Deletions S:C136-, S:N137-, S:D138-, S:P139-, S:F140-, S:L141-, S:G142-, S:V143-, S:Y144-	Deletions S:C136-, S:N137-, S:D138-, S:P139-, S:F140-, S:L141-, S:G142-, S:V143-, S:Y144-
S:R190S	S:R190S
S:I210T	-
-	S:D215H
S:R346S	S:R346S
S:N394S	S:N394S
S:Y449N	S:Y449N
-	S:E484K
S:F490R	S:F490S
S:N501Y	S:N501Y
S:D614G	S:D614G
S:P681H	S:P681H
S:T859N	S:T859N
S:D936H	-
-	S:D1139H

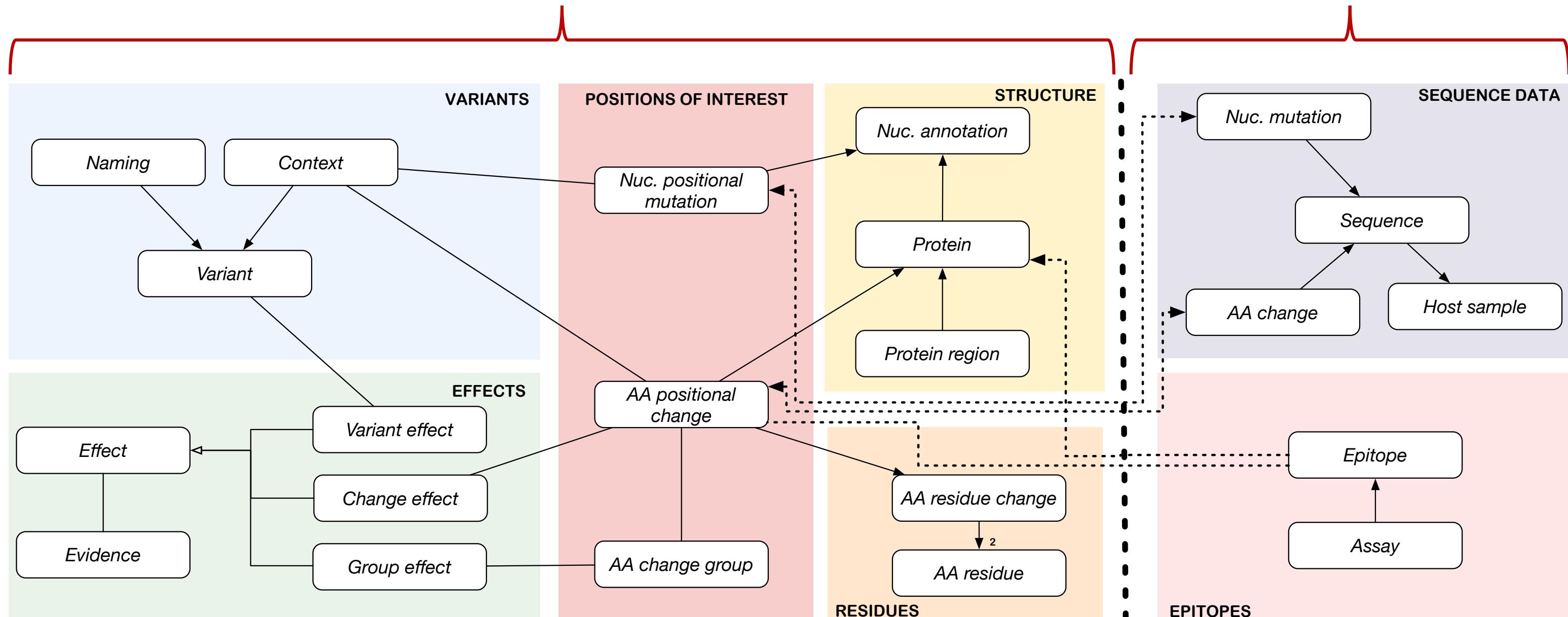
---

# Knowledge Tools

# CoV2K knowledge base abstract model

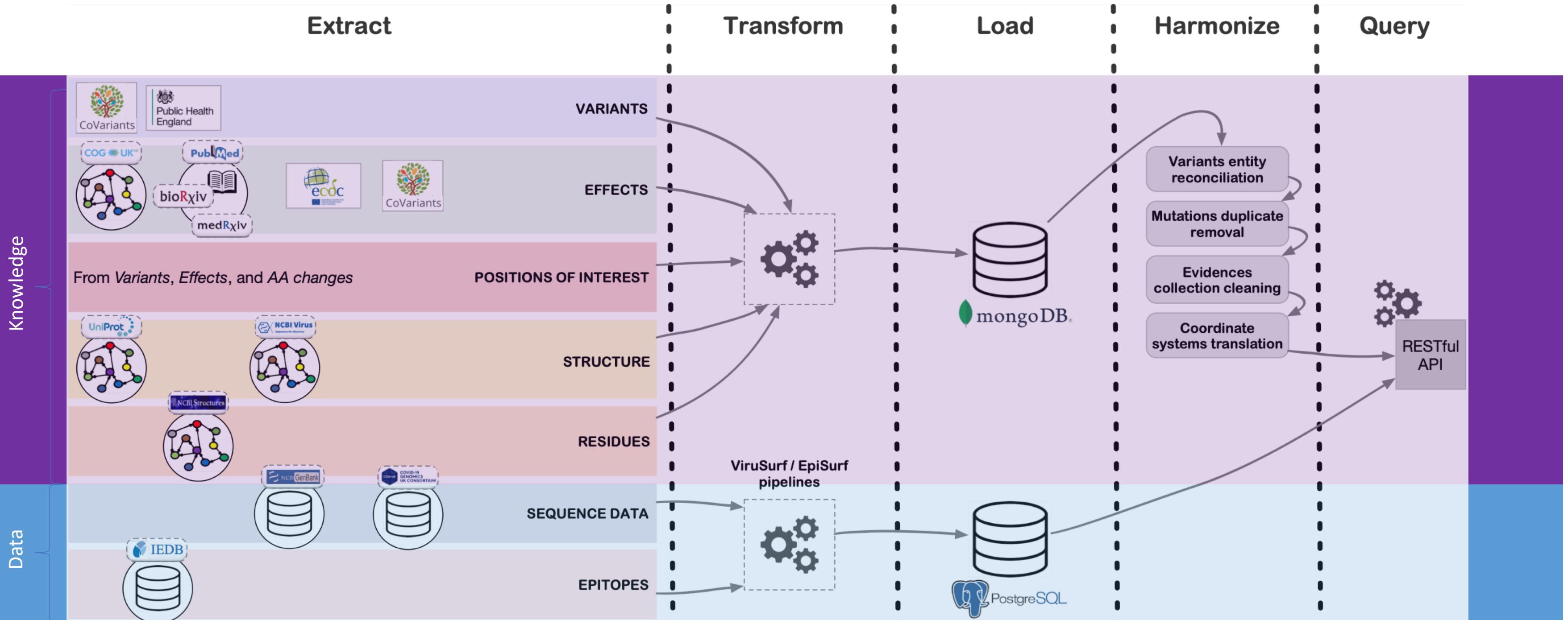


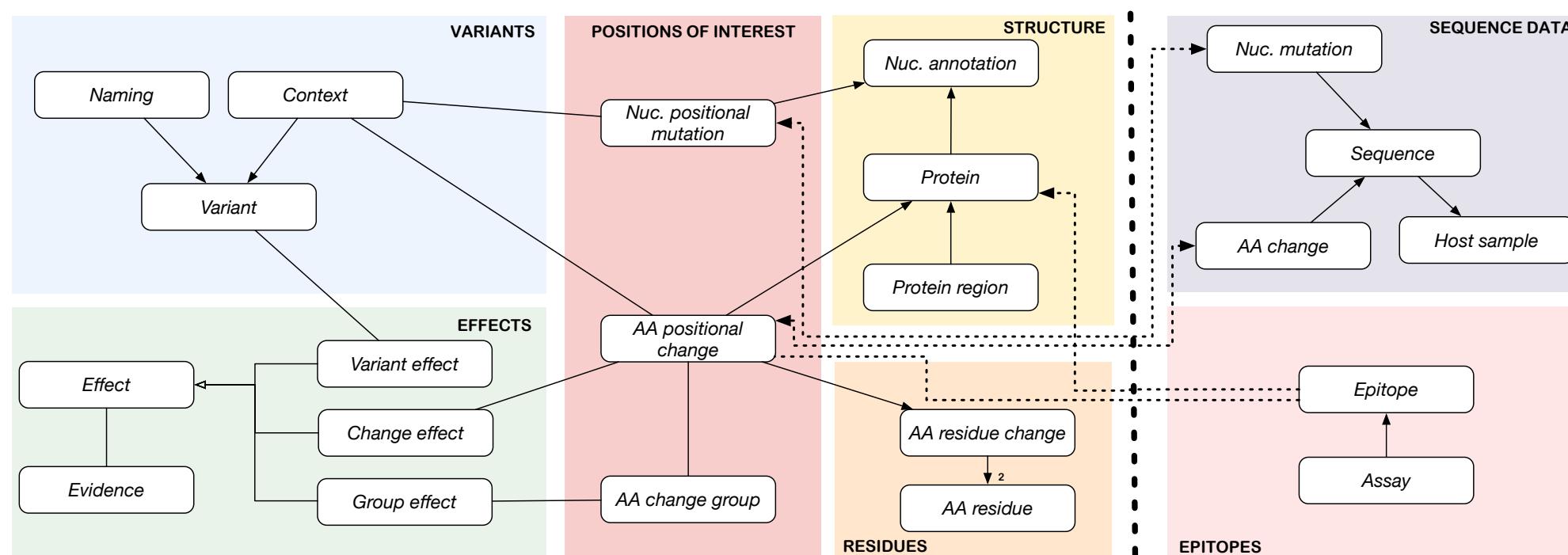
## Knowledge representation



Alfonsi, T., Al Khalaf, R., Ceri, S., & Bernasconi, A. CoV2K model, a comprehensive representation of SARS-CoV-2 knowledge and data interplay. Submitted to Scientific Data Journal – 2-year IF 2019: 5.5, SJR: Q1

# CoV2K knowledge and data integration pipeline





## One endpoint for each entity of CoV2K:

- Without parameters;
- With a *same entity identifier* as a path parameter (returning only one instance);
- With an *attribute-value pair* as a query parameter (filter on the entity);
- With *another entity identifier* query parameter (returning the set of instances of the first entity that are linked to the instances of the identified second entity)

Possibility to traverse paths of the graph  
(with the «combine» endpoints)

## Example queries:

- What are the characteristics (Grantham distance and type) of the residue changes of the Alpha variant?
- Which amino acid changes of VOC-20DEC-02 fall within the Receptor Binding Domain (RBD)?
- Which are the effects of the variants that include the Spike amino acid change D614G?

# Collection of effects of Omicron mutations

---



Mutation: S:Y145D

LOWER sensitivity to neutralizing mabs, ['Haslwanter et al, <https://doi.org/10.1101/2021.06.10.447999>']

Mutation: S:K417N

LOWER sensitivity to vaccine sera, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

LOWER sensitivity to vaccine sera, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>']

LOWER sensitivity to vaccine sera, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Chen et al.(2021), <https://doi.org/10.1038/s41591-021-01294-w>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Chen et al.(2021), <https://doi.org/10.1038/s41591-021-01294-w>']

HIGHER binding to host receptor, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER protein stability, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER protein flexibility, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER binding to abs, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

UNKNOWN binding to host receptor, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

Mutation: S:N440K

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Rappazzo et al. (2021), <https://doi.org/10.1126/science.abf4830>']

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>', 'Wang et al. (2021), <https://doi.org/10.1101/2021.01.15.426911>', 'Rappazzo et al. (2021), <https://doi.org/10.1126/science.abf4830>', 'Weisblum et al. (2020), <https://doi.org/10.7554/eLife.61312>']

LOWER protein stability, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER protein flexibility, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER binding to abs, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

Mutation: S:G446S

LOWER sensitivity to neutralizing mabs, ['Starr et al. (2020), <https://doi.org/10.1101/2020.11.30.405472>']

Mutation: S:S477N

LOWER sensitivity to neutralizing mabs, ['Liu et al. (2020), <https://doi.org/10.1016/j.chom.2021.01.014>']

HIGHER binding to host receptor, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER infectivity, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER viral transmission, ['Chen et al. (2020), <https://doi.org/10.1016/j.jmb.2020.07.009>']

HIGHER protein stability, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER protein flexibility, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

LOWER binding to abs, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER binding to host receptor, ['Singh et al. (2021), <https://doi.org/10.3390/v13030439>']

HIGHER binding to host receptor, ['Singh et al. (2021), <https://doi.org/10.1038/s41598-021-83761-5>']