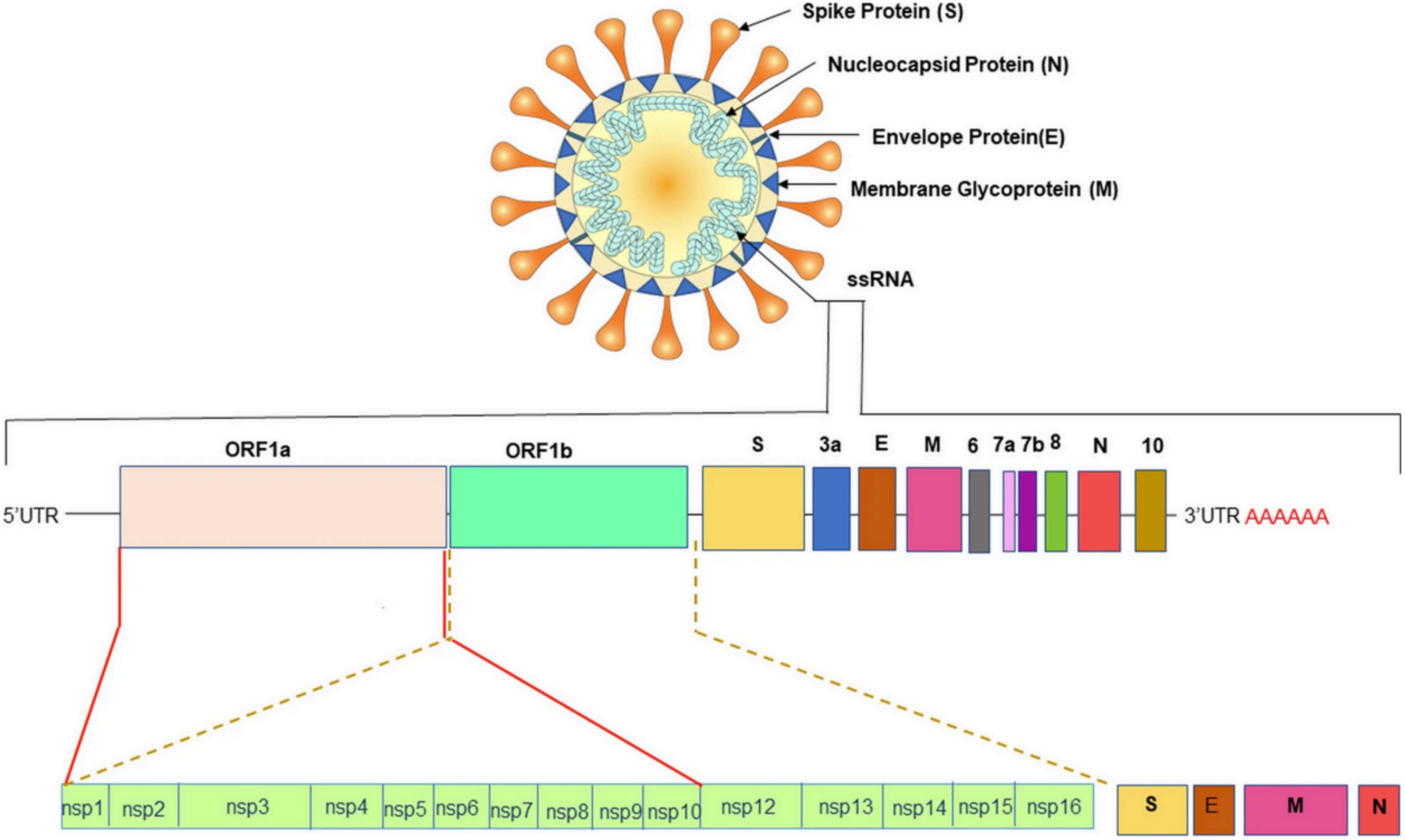
The evolutionary history of Coronaviruses provides insights for the COVID-19 pandemic and the future evolutionary paths of SARS-CoV-2

Grigorios D. Amoutzias
Associate Professor of Bioinformatics with emphasis on Microbiology
Department of Biochemistry and Biotechnology
University of Thessaly, Larisa
Greece

Coronaviruses

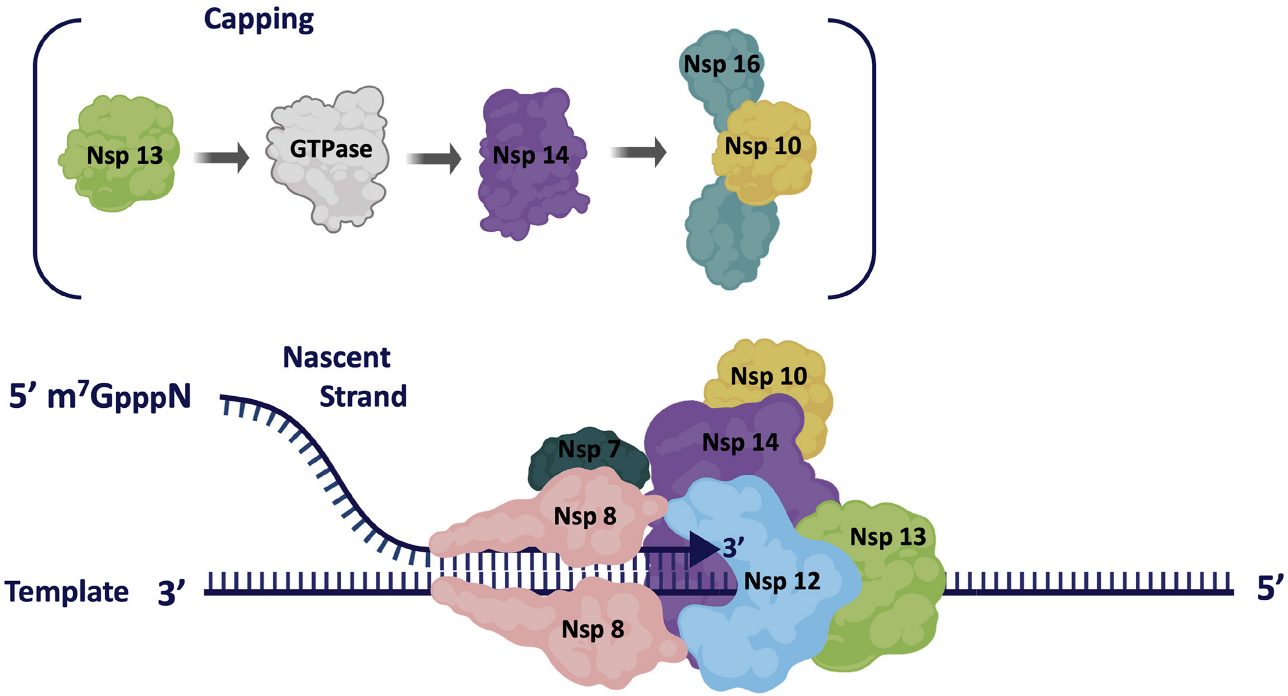
- Enveloped, spherical, positive-sense single-stranded RNA viruses.
- Very large genome size: 25-32Kb.
- 5’UTR, at least 6 core ORFs found in all CoVs, a highly variable number of accessory ORFs and a 3’UTR.



From: SARS coronavirus 2: from genome to infectome

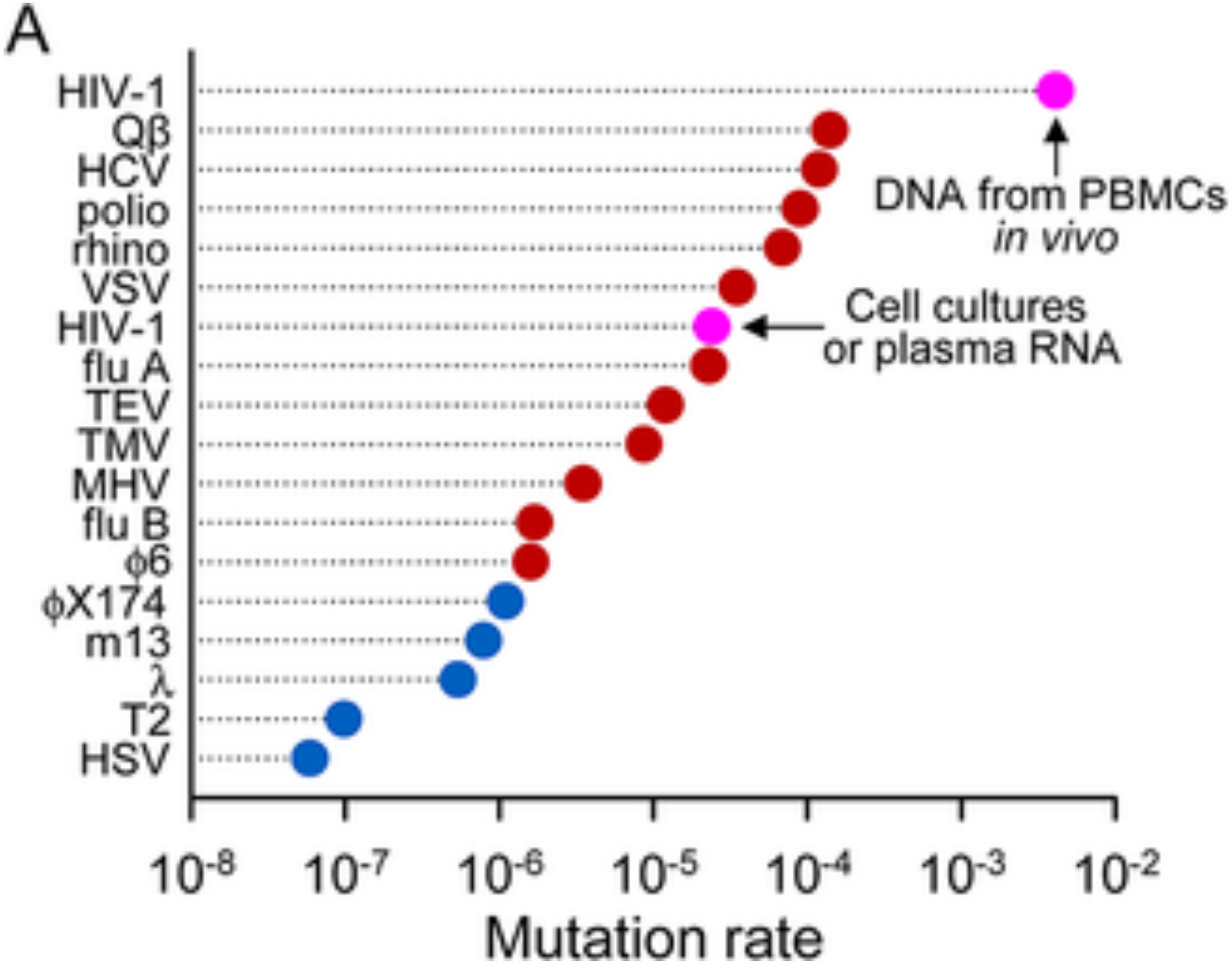
Nsp14 exonuclease

Replication proof-reading mechanism lowers mutation rate and allows for large genomes.



Robson et al., 2020; Cell

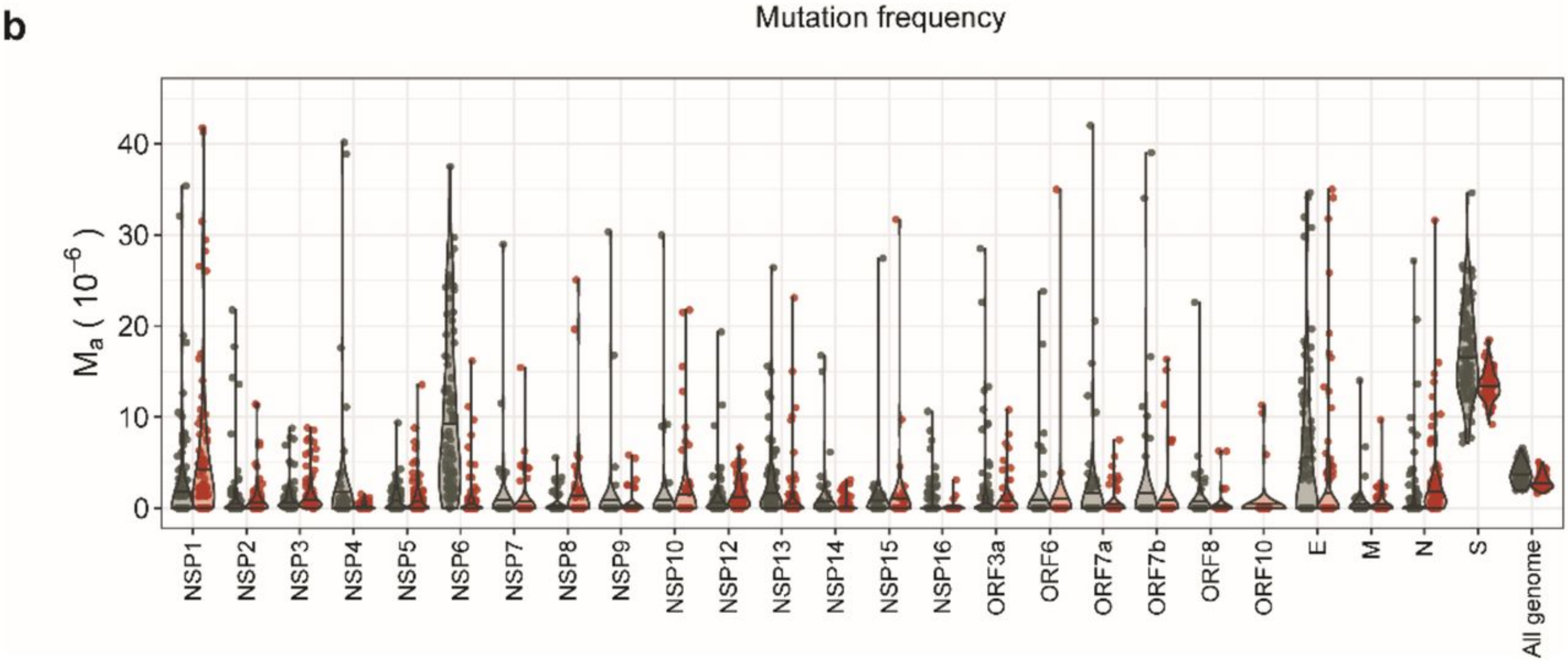
Mutation rates of viruses



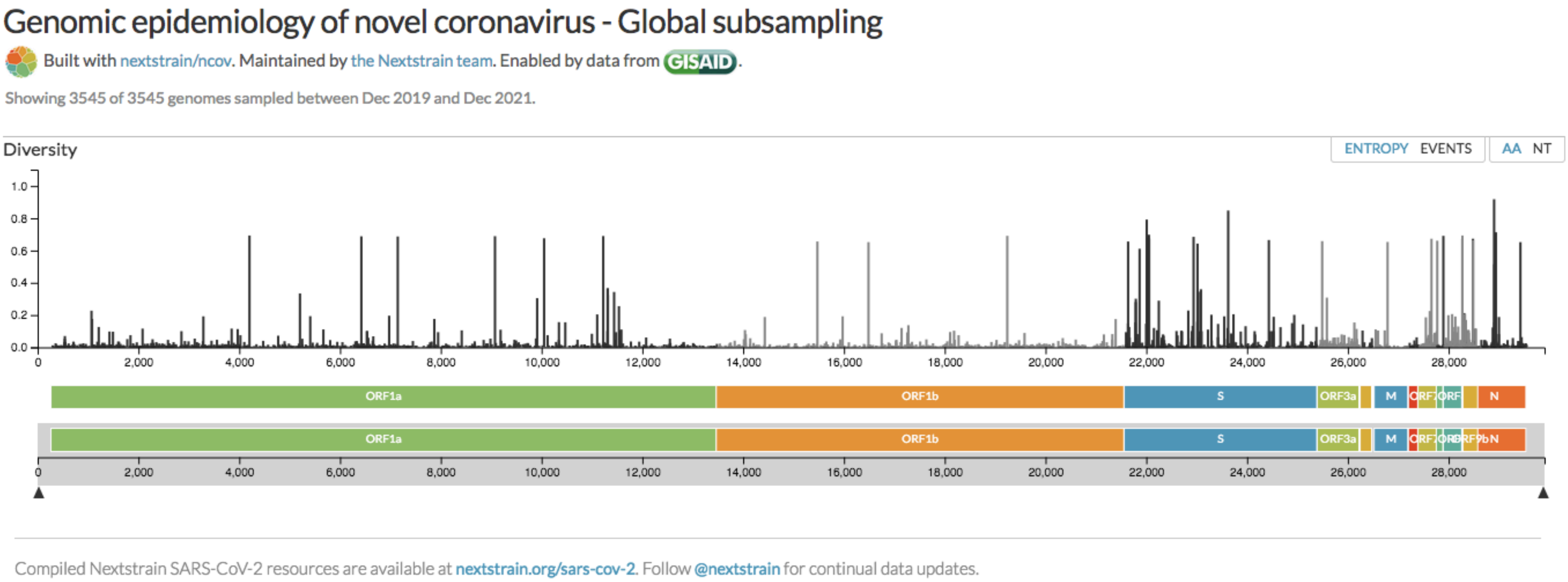
Cuevas et al., 2015; Plos Biology

Evolution by point mutations

- A background mutation rate of 2.9-3.7×10^{-6}/nt/replication cycle for SARS-CoV-2.
- The Spike gene has a mutation rate at least 4-5 times higher than the rest of the genome.
- Omicron (B1.1.529): 30/50 total mutations in Spike



Amicone et al., 2021; BioRxiv



Genomic epidemiology of novel coronavirus - Global subsampling

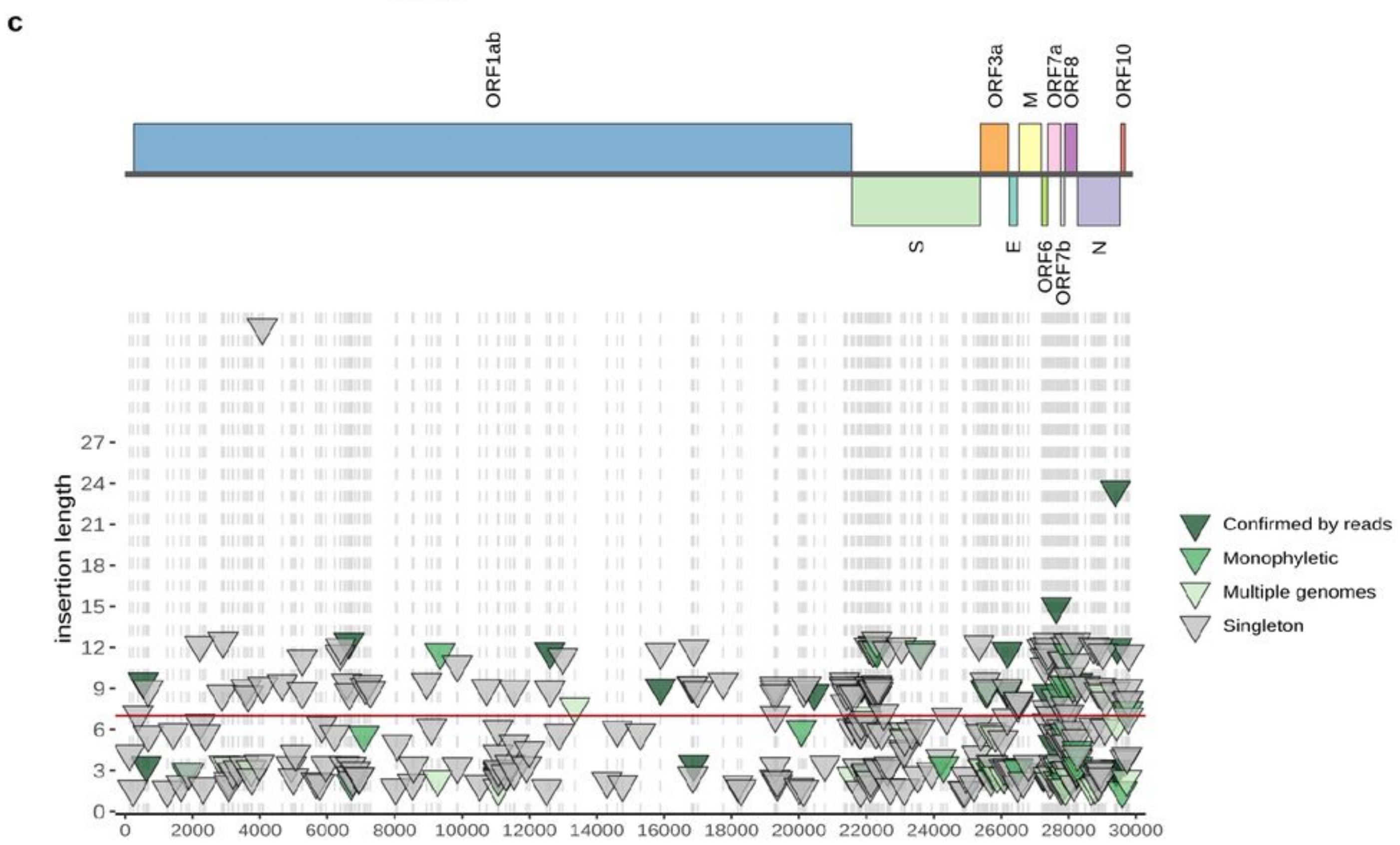
Built with nextstrain/ncov. Maintained by the Nextstrain team. Enabled by data from GISAID.

Showing 3545 of 3545 genomes sampled between Dec 2019 and Dec 2021.

Diversity

Compiled Nextstrain SARS-CoV-2 resources are available at nextstrain.org/sars-cov-2. Follow @nextstrain for continual data updates.

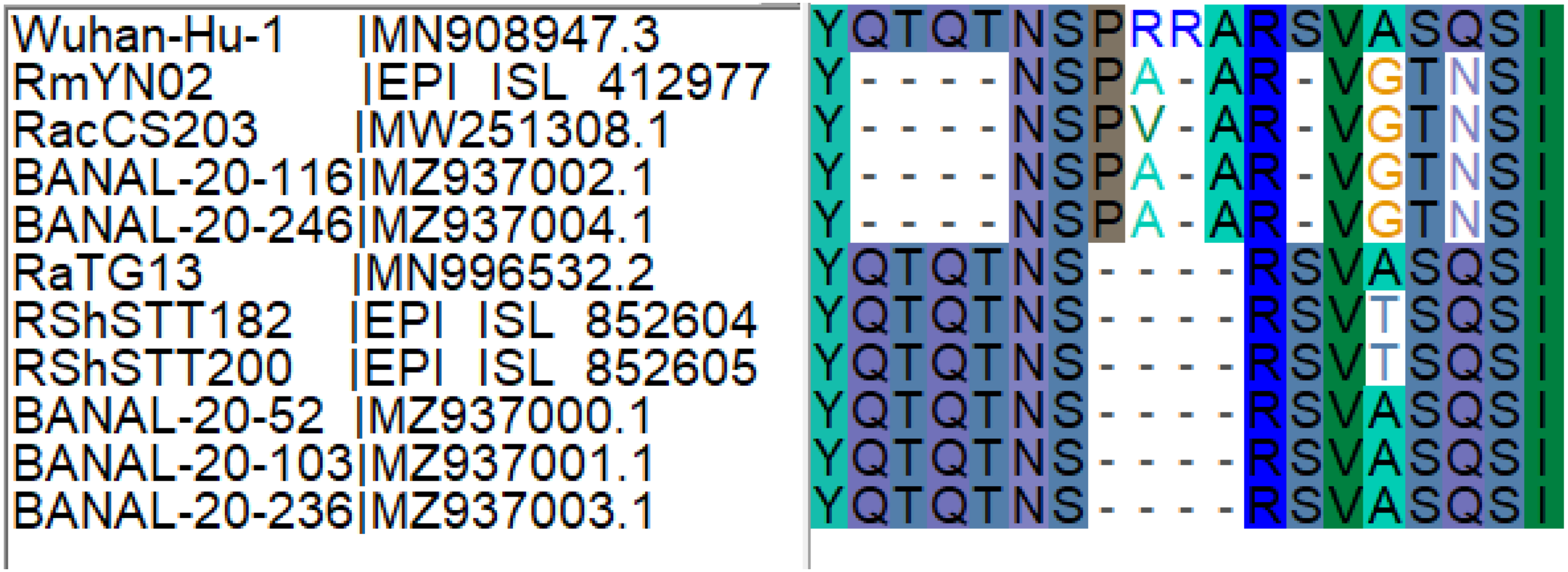
Evolution by insertions/deletions



Insertions in SARS-CoV-2 genomes. Distribution of inserts along the genome. Each triangle represents one insertion event. The level of confidence in each variant is represented by color: dark green, confirmed by sequencing read analysis; green, monophyletic in the tree, no read data available; light green, observed multiple times, but not monophyletic; grey, singletons (Supplementary Table 2). The positions of inserts are marked with grey dashed lines.

Garushyants et al., 2021; BioRxiv

Furin Cleavage Site (FCS)



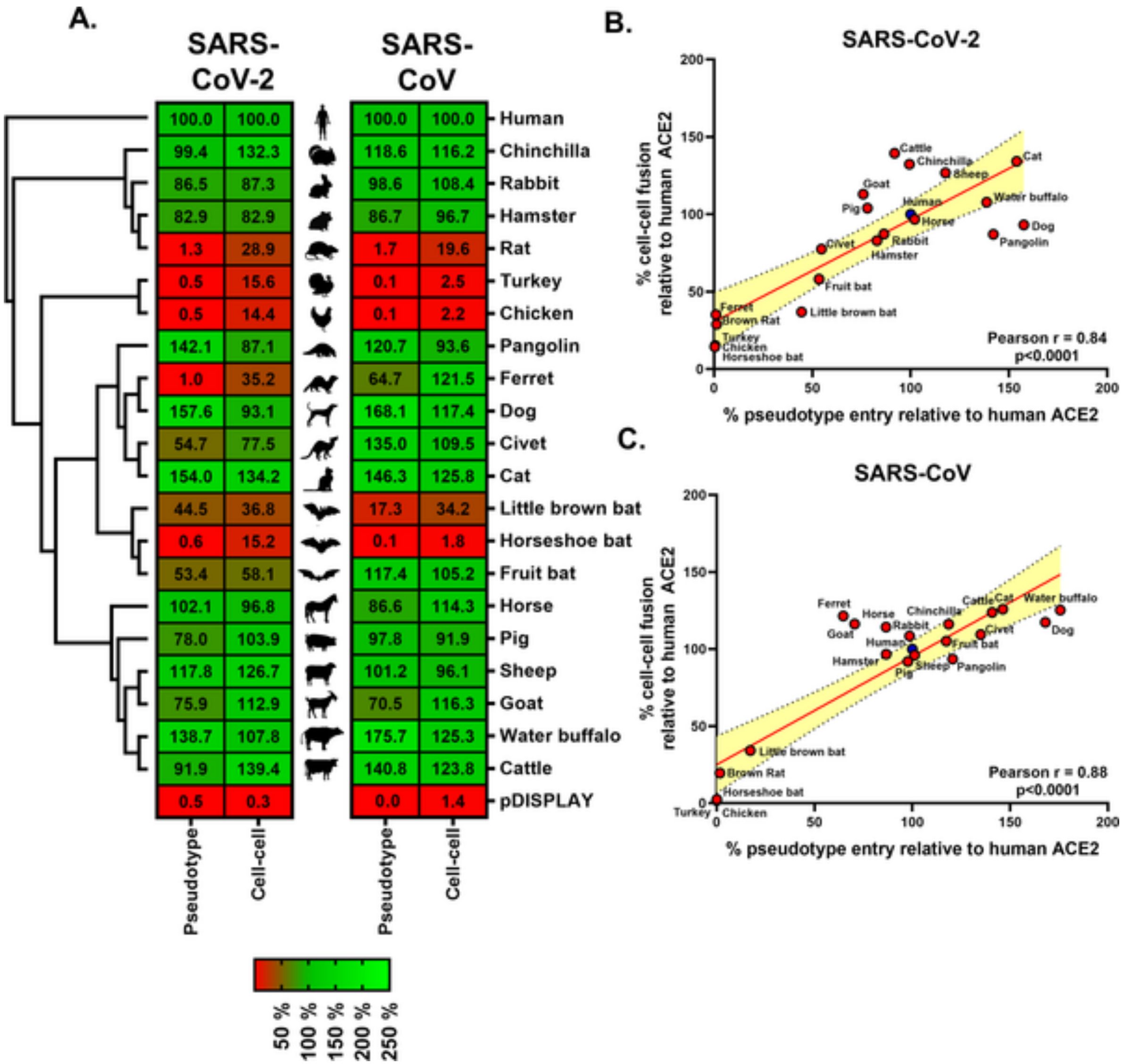
Wuhan-Hu-1	|MN908947.3	YQTQTNSPRRARSVASQSI
RmYN02	|EPI ISL 412977	Y----NSPA-AR-VGTNSI
RacCS203	|MW251308.1	Y----NSPV-AR-VGTNSI
BANAL-20-116	|MZ937002.1	Y----NSPA-AR-VGTNSI
BANAL-20-246	|MZ937004.1	Y----NSPA-AR-VGTNSI
RaTG13	|MN996532.2	YQTQTNS----RSVASQSI
RShSTT182	|EPI ISL 852604	YQTQTNS----RSVTSQSI
RShSTT200	|EPI ISL 852605	YQTQTNS----RSVTSQSI
BANAL-20-52	|MZ937000.1	YQTQTNS----RSVASQSI
BANAL-20-103	|MZ937001.1	YQTQTNS----RSVASQSI
BANAL-20-236	|MZ937003.1	YQTQTNS----RSVASQSI

Lytras, 2021; Virological.org

Similar inserts are found in other Sarbecoviruses.
Omicron variant bears two mutations (N679K and P681H) at the FCS

The progenitor of SARS-CoV-2 was already a generalist virus that did not need many mutations to adapt to its human hosts.

The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins



	SARS-CoV-2 Pseudotype	SARS-CoV-2 Cell-cell	SARS-CoV Pseudotype	SARS-CoV Cell-cell
Human	100.0	100.0	100.0	100.0
Chinchilla	99.4	132.3	118.6	116.2
Rabbit	86.5	87.3	98.6	108.4
Hamster	82.9	82.9	86.7	96.7
Rat	1.3	28.9	1.7	19.6
Turkey	0.5	15.6	0.1	2.5
Chicken	0.5	14.4	0.1	2.2
Pangolin	142.1	87.1	120.7	93.6
Ferret	1.0	35.2	64.7	121.5
Dog	157.6	93.1	168.1	117.4
Civet	54.7	77.5	135.0	109.5
Cat	154.0	134.2	146.3	125.8
Little brown bat	44.5	36.8	17.3	34.2
Horseshoe bat	0.6	15.2	0.1	1.8
Fruit bat	53.4	58.1	117.4	105.2
Horse	102.1	96.8	86.6	114.3
Pig	78.0	103.9	97.8	91.9
Sheep	117.8	126.7	101.2	96.1
Goat	75.9	112.9	70.5	116.3
Water buffalo	138.7	107.8	175.7	125.3
Cattle	91.9	139.4	140.8	123.8
pDISPLAY	0.5	0.3	0.0	1.4

Conceicao et al., 2021; Plos Biology

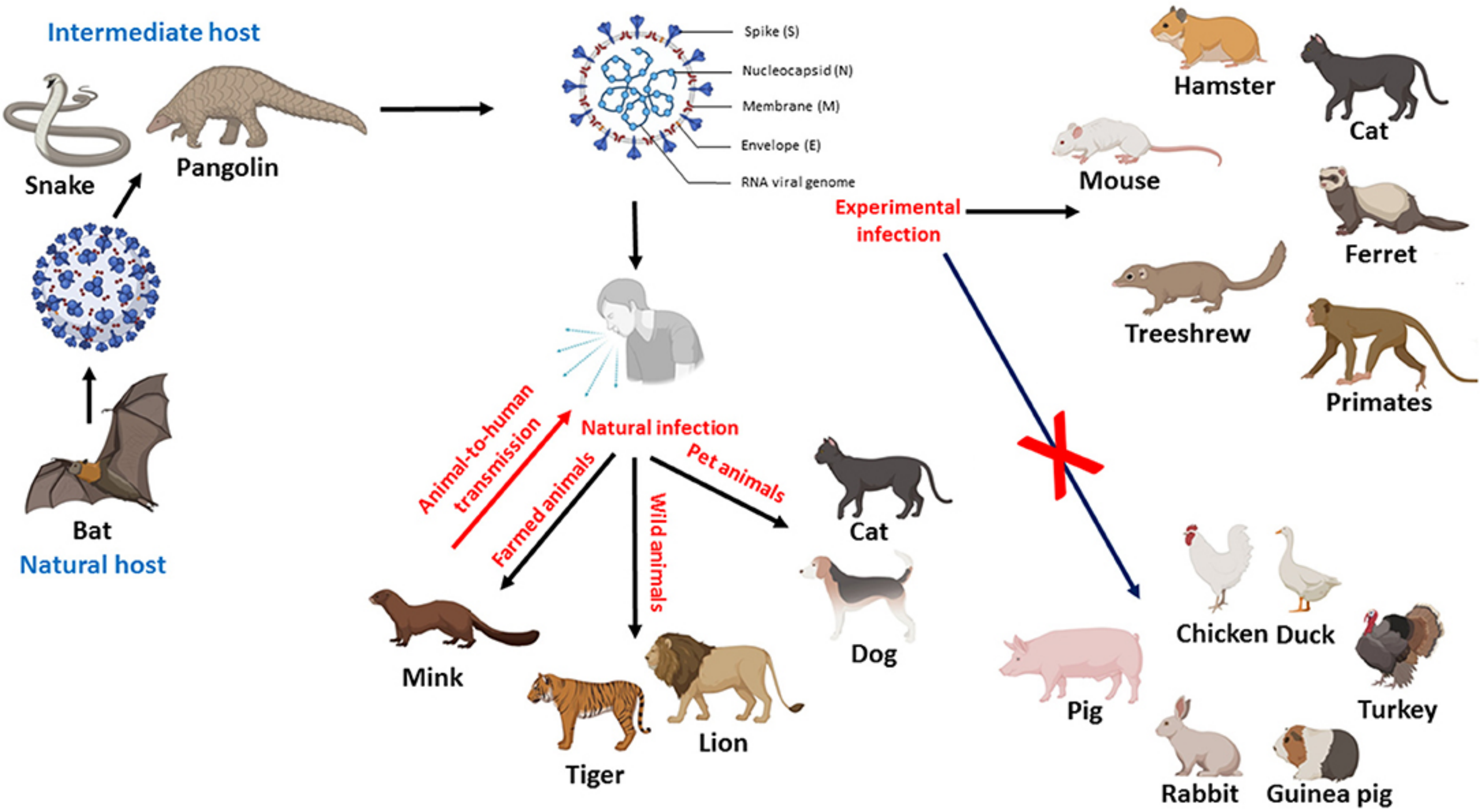
Many animals can be infected by SARS-CoV-2.

SARS-CoV-2 did not undergo significant positive selection during the first year of the pandemic.

Laos bat SARS-CoV-2-like virus can easily bind to hACE2.

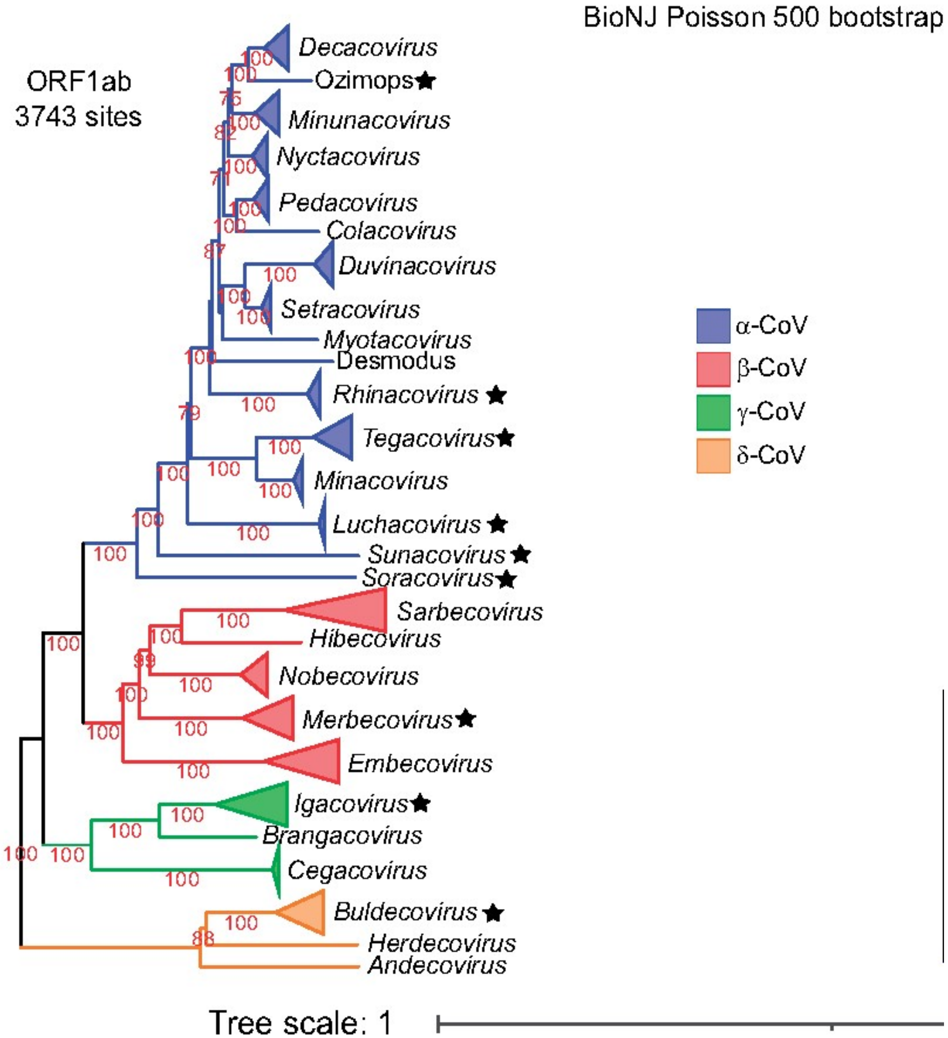
The progenitor of SARS-CoV-2 was already a generalist virus that did not need many mutations to adapt to its human hosts.

The SARS-CoV-2 Spike protein has a broad tropism for mammalian ACE2 proteins



Mahdy et al., 2021; Frontiers in Veterinary Sciences

Coronavirus classification



Stars indicate the CoV subgenera that are implicated in recombinat

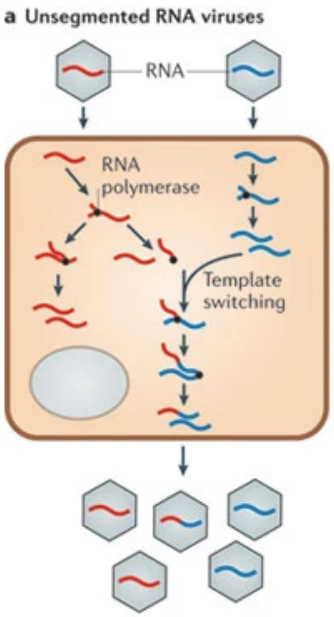
ICTV classification:

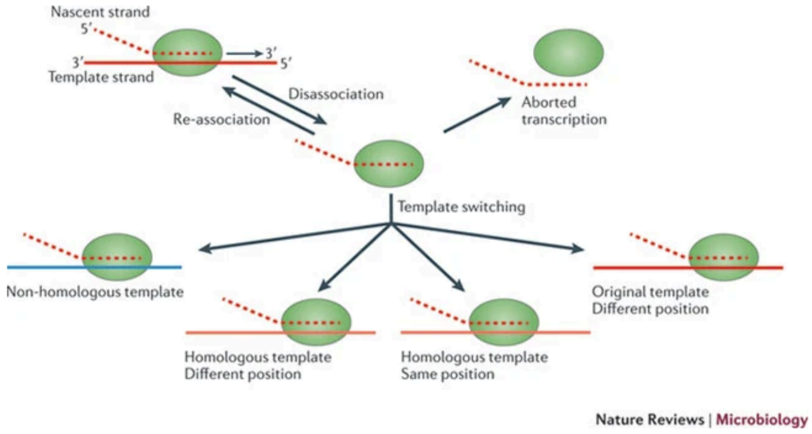
- 4 genera
- 25 subgenera

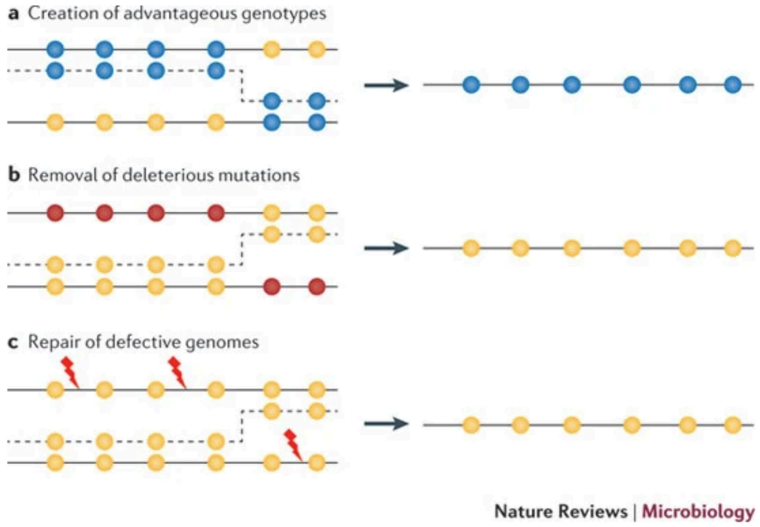
Nikolaidis et al., 2021; Mol. Biol. Evol

Evolution by recombination

RNA viruses recombine frequently by template switching





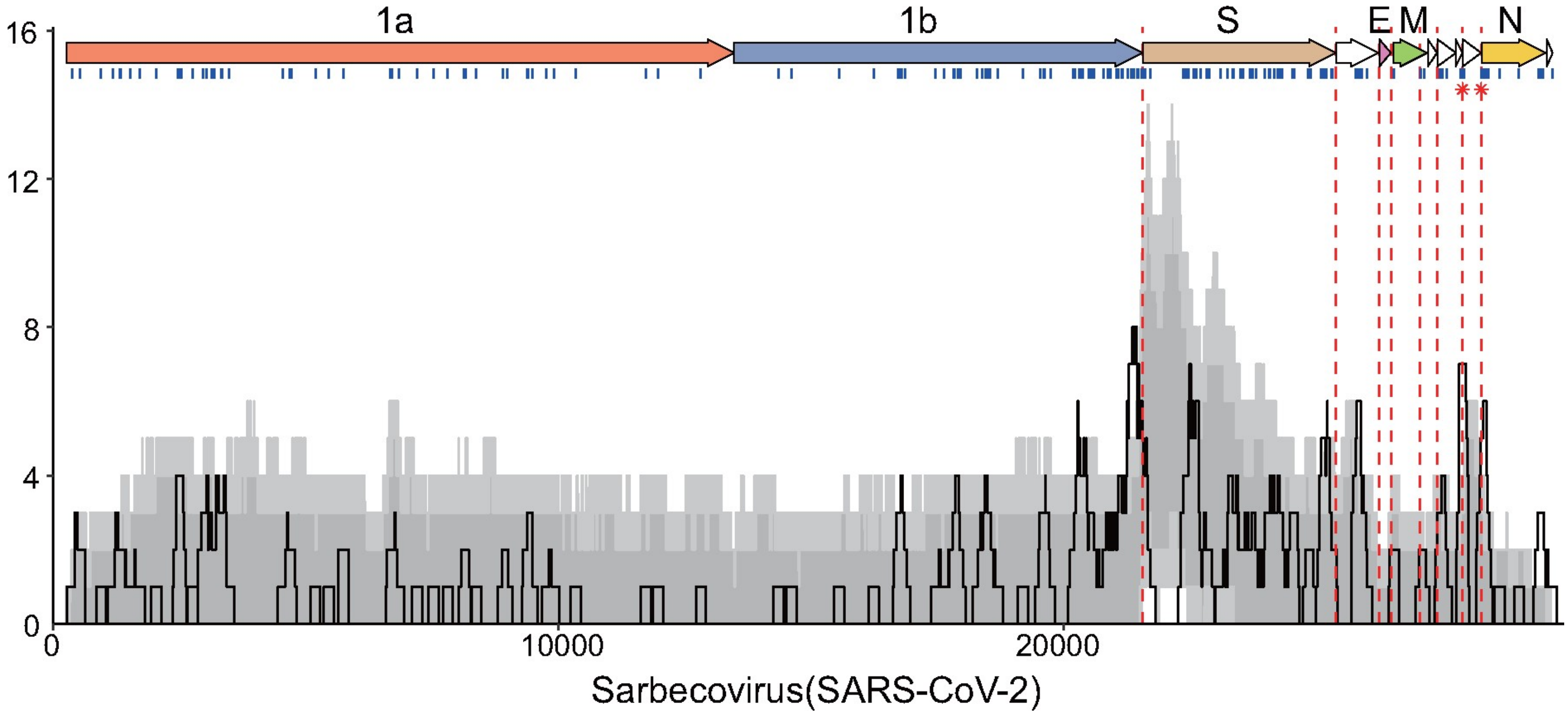


Types of recombination:

- Intratypic homologous
- Intertypic homologous
- Non-homologous

Simon-Loriere and Holmes., 2021; Nature Reviews Microbiology

Intratypic homologous recombination

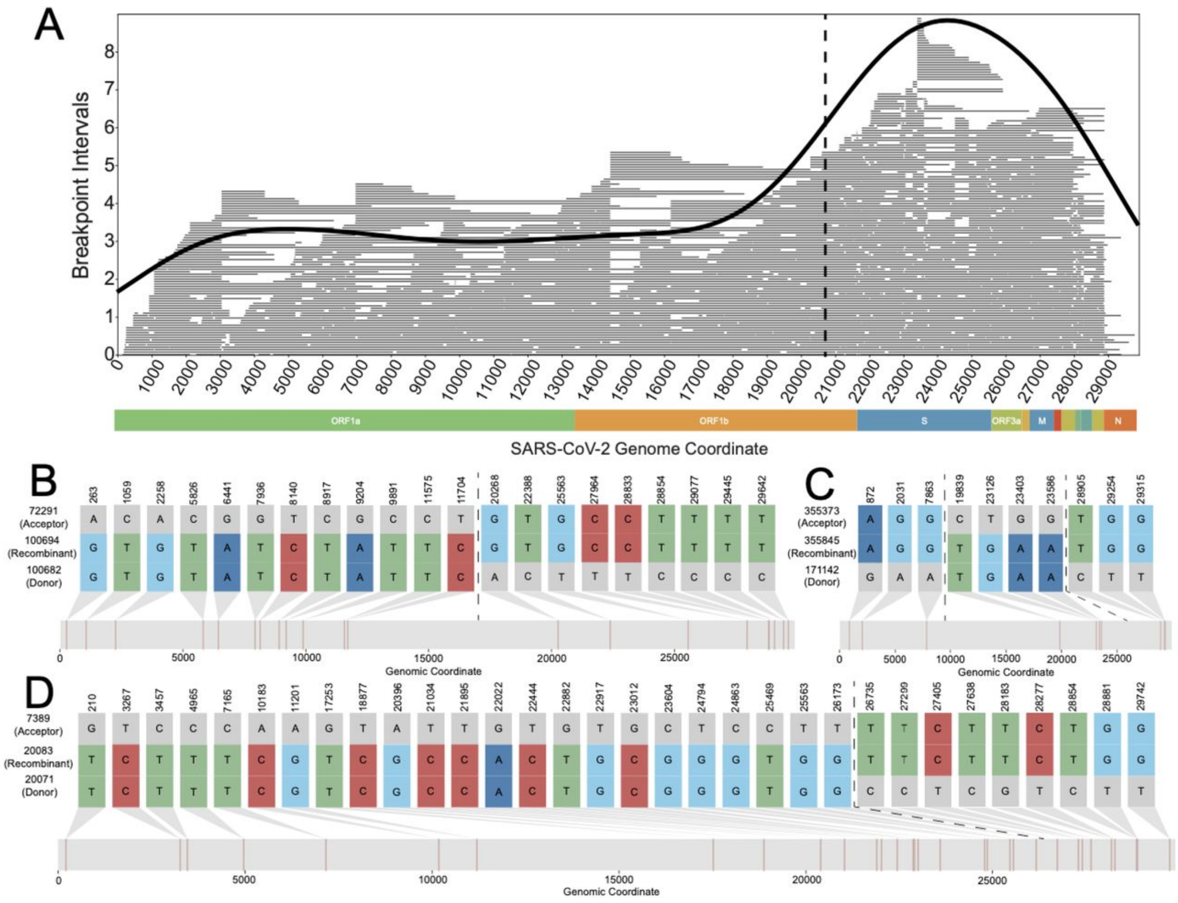


Very frequent.
Many analyses for different CoVs have observed many such events. Spike is a hotspot.

973 intratypic recombination events within 16 different subgenera.
Many of them localized around transcriptional start sites and especially around the Spike.

Yang et al., 2021; Mol Biol Evol

Intra-SARS-CoV-2 homologous recombination



B

	263	1059	2258	5826	6441	7936	8140	8917	9204	9891	11575	11704	20268	22388	25563	27964	28833	28854	29077	29445	29642
72291 (Acceptor)	A	C	A	C	G	G	T	C	G	C	C	T	G	T	G	C	C	T	T	T	T
100694 (Recombinant)	G	T	G	T	A	T	C	T	A	T	T	C	G	T	G	C	C	T	T	T	T
100682 (Donor)	G	T	G	T	A	T	C	T	A	T	T	C	A	C	T	T	T	C	C	C	C

C

	872	2031	7863	19839	23126	23403	23586	28905	29254	29315
355373 (Acceptor)	A	G	G	C	T	G	G	T	G	G
355845 (Recombinant)	A	G	G	T	G	A	A	T	G	G
171142 (Donor)	G	A	A	T	G	A	A	C	T	T

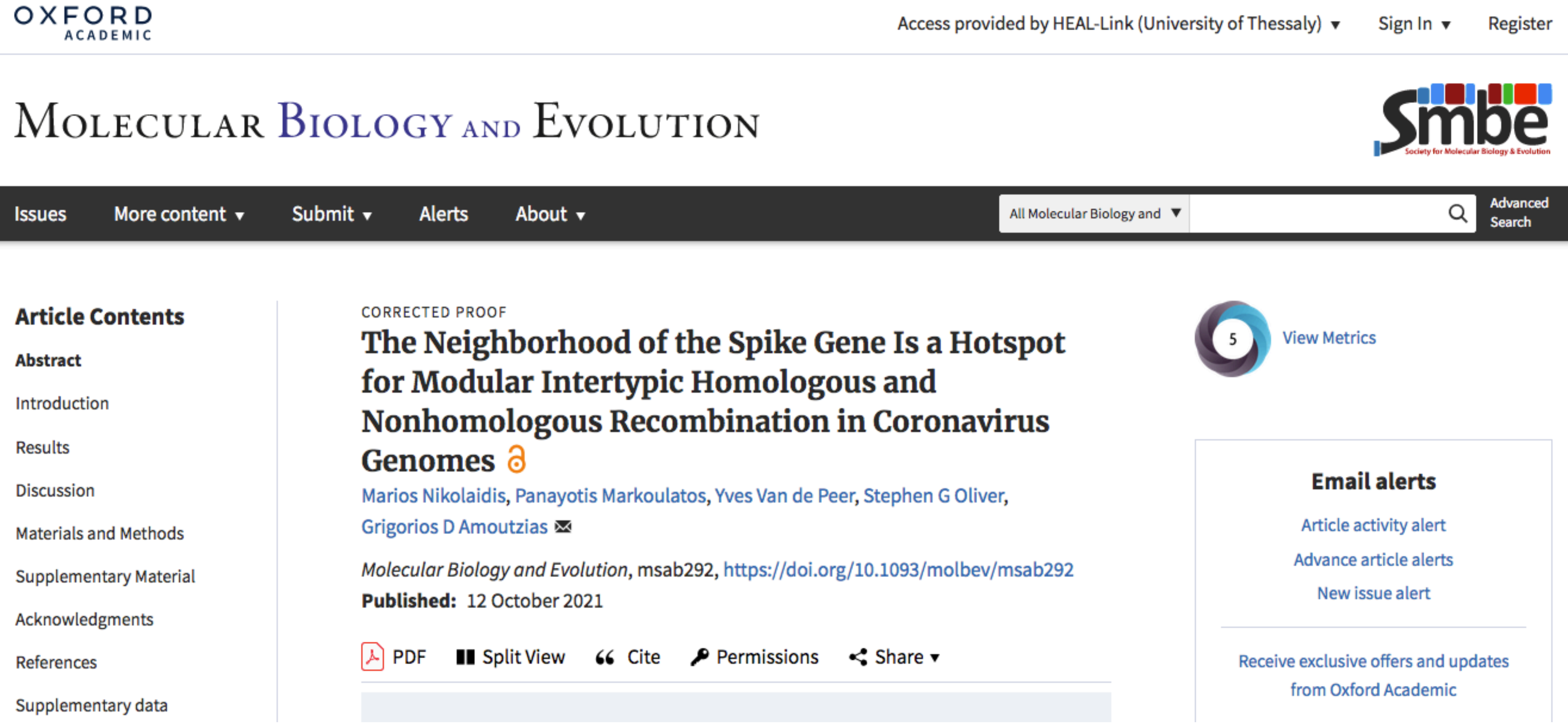
D

	210	3267	3457	4965	7165	10183	11201	17253	18877	20396	21034	21895	22022	22444	22882	22917	23012	23604	24794	24863	25469	25563	26173	26735	27299	27405	27638	28183	28277	28854	28881	29742
7389 (Acceptor)	G	T	C	C	C	A	A	G	T	A	T	T	G	T	G	T	G	C	T	C	C	T	T	T	T	C	T	T	C	T	G	G
20083 (Recombinant)	T	C	T	T	T	C	G	T	C	G	C	C	A	C	T	G	C	G	G	G	T	G	G	T	T	C	T	T	C	T	G	G
20071 (Donor)	T	C	T	T	T	C	G	T	C	G	C	C	A	C	T	G	C	G	G	G	T	G	G	C	C	T	C	G	T	C	T	T

Analysis of 1.6 million genomes. 2.7% are recombinant. Many recombination events localize at the Spike.

Turakhia et al., 2021; Biorxiv

CoV intertypic homologous recombination



CORRECTED PROOF

The Neighborhood of the Spike Gene Is a Hotspot for Modular Intertypic Homologous and Nonhomologous Recombination in Coronavirus Genomes

Marios Nikolaidis, Panayotis Markoulatos, Yves Van de Peer, Stephen G Oliver, Grigorios D Amoutzias

Molecular Biology and Evolution, msab292, https://doi.org/10.1093/molbev/msab292

Published: 12 October 2021

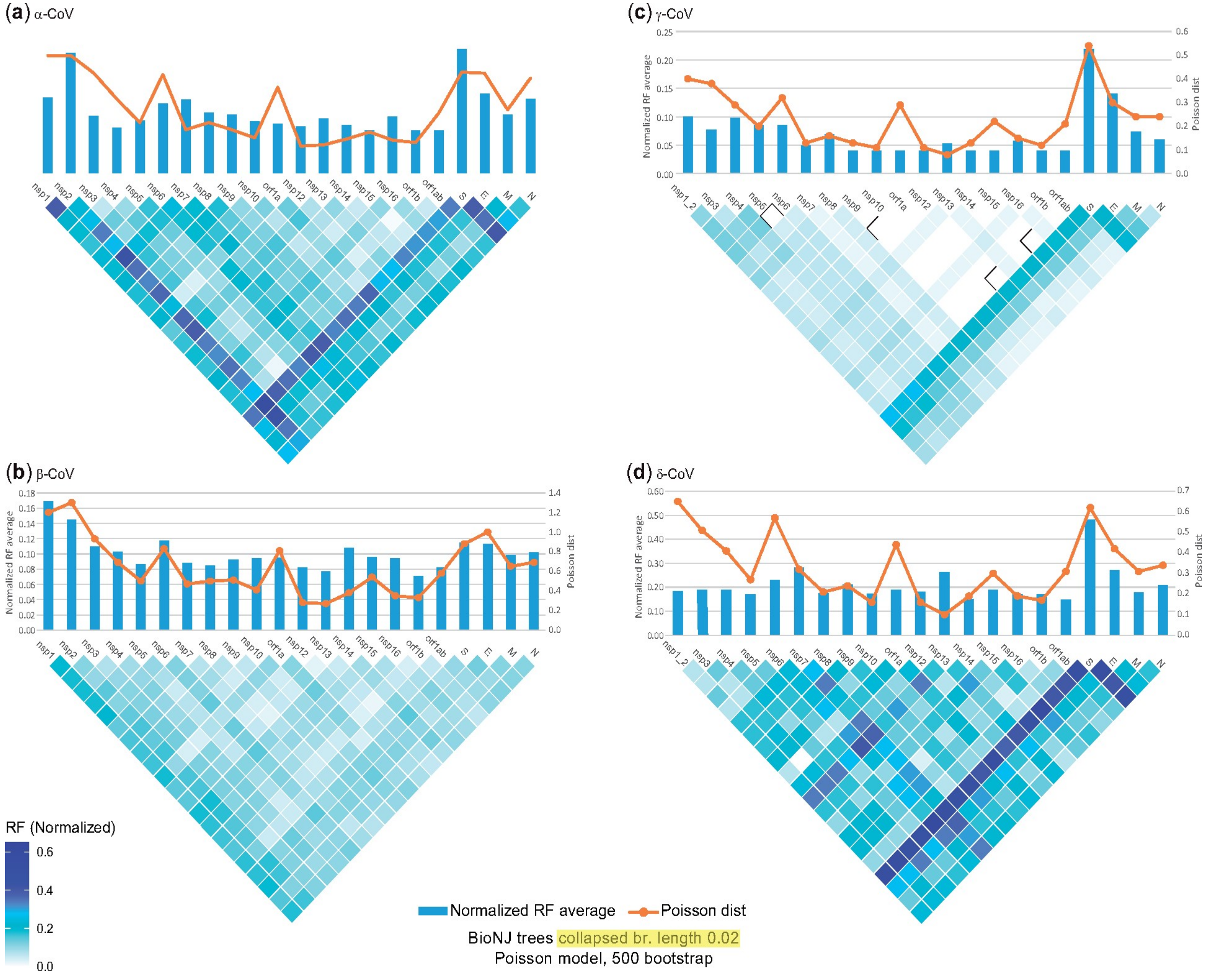
Intertypic CoV homologous recombination

Analyzed recombination events among genomes of different subgenera by:

- Phylogenetic incongruence
  - Robinson-Foulds Metric of incongruence
  - Tanglegrams
- Similarity Plots
- Shimodaira-Hasegawa test

Nikolaidis et al., 2021; Mol Biol Evol

Robinson-Foulds Metric of incongruence



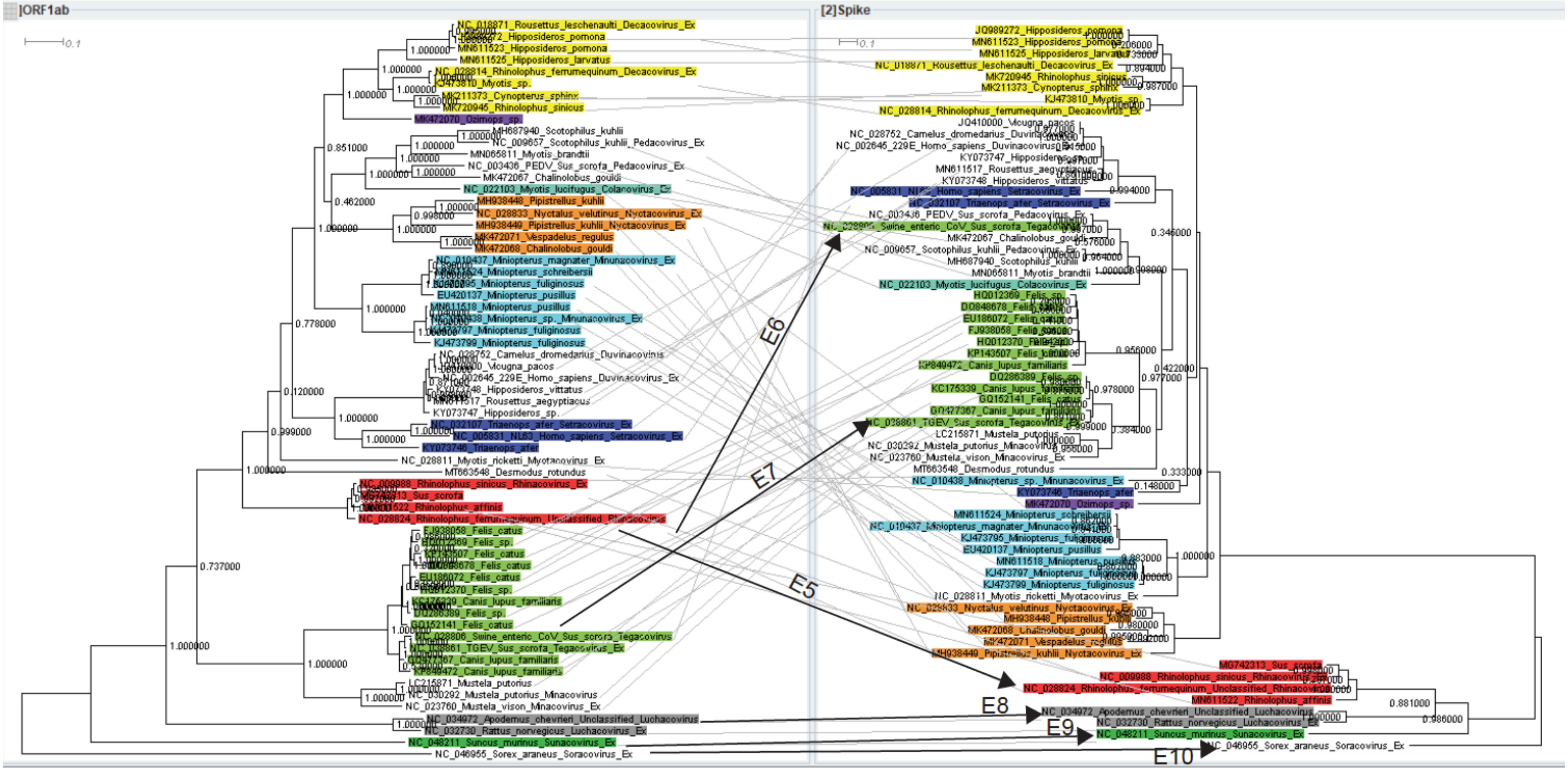
BioNJ trees collapsed br. length 0.02
Poisson model, 500 bootstrap

Incongruence among different core regions for of the four CoV genera (A–D) based on the normalized RF method. The darker the color, the higher the incongruence

Nikolaidis et al., 2021; Mol Biol Evol

Incongruence by tanglegrams

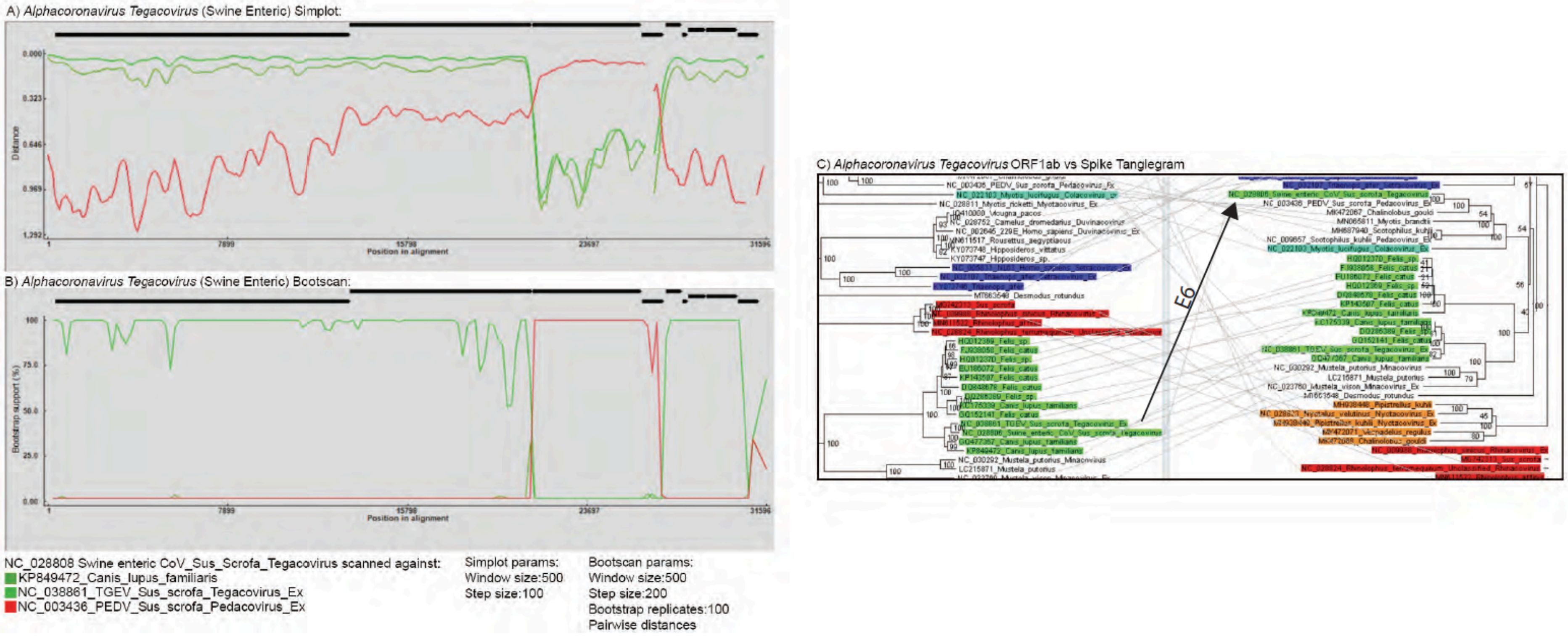
Supp. fig. 21: *Alphacoronavirus* ORF1ab - Spike PhyML Tanglegram (Events 5-10)



Nikolaidis et al., 2021; Mol Biol Evol

Incongruence by tanglegrams, verified by Similarity Plots

Supp. fig. 24: Recombination event 6 Similarity plot and Bootscan analyses



NC_028808 Swine enteric CoV_Sus_Scrofa_Tegacovirus scanned against:

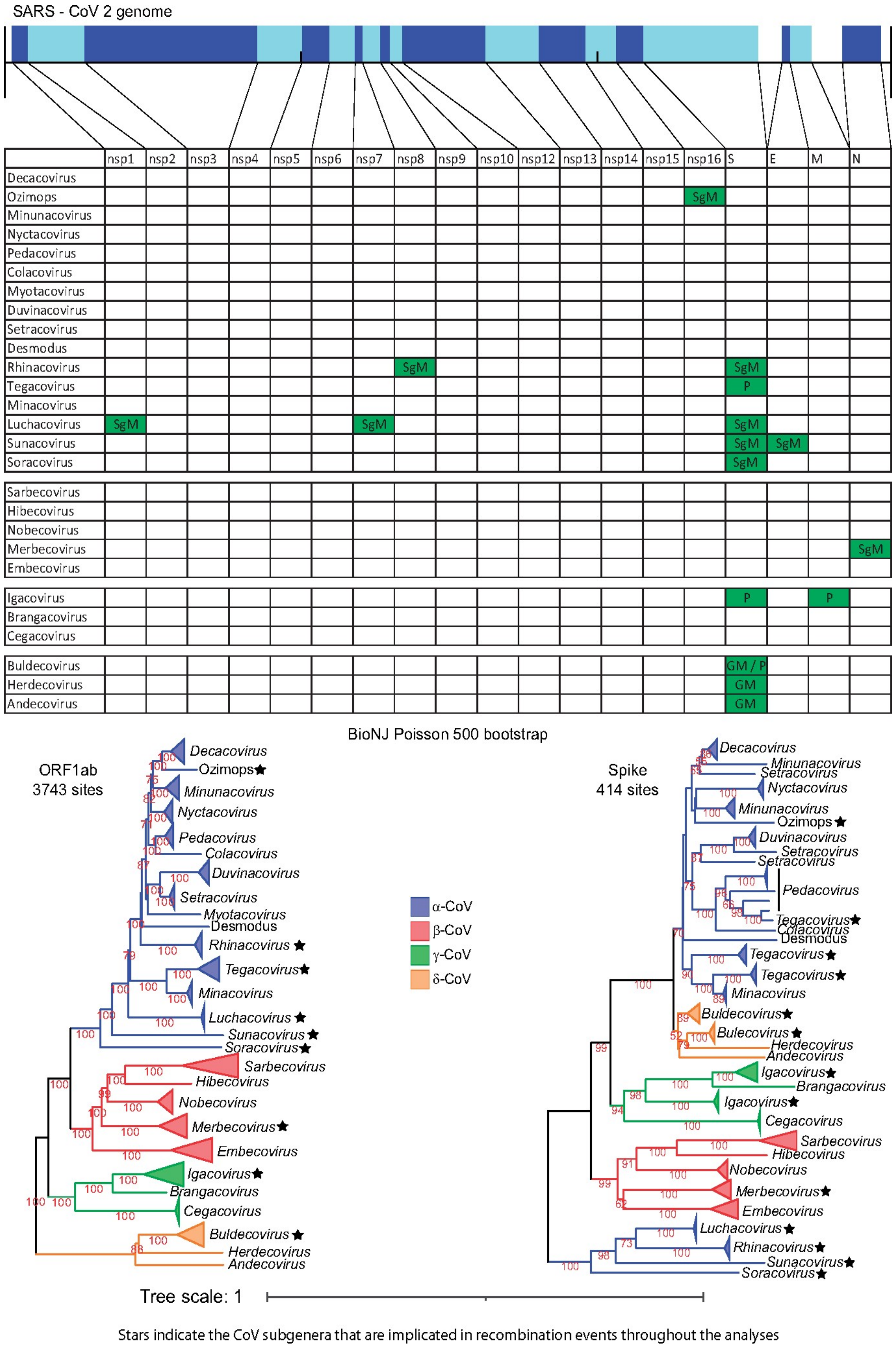
- 🟩 KP849472_Canis_lupus_familiaris
- 🟩 NC_038861_TGEV_Sus_scrofa_Tegacovirus_Ex
- 🟥 NC_003436_PEDV_Sus_scrofa_Pedacovirus_Ex

Simplot params:
Window size:500
Step size:100

Bootscan params:
Window size:500
Step size:200
Bootstrap replicates:100
Pairwise distances

Nikolaidis et al., 2021; Mol Biol Evol

Incongruence by tanglegrams



	nsp1	nsp2	nsp3	nsp4	nsp5	nsp6	nsp7	nsp8	nsp9	nsp10	nsp12	nsp13	nsp14	nsp15	nsp16	S	E	M	N
Decacovirus																			
Ozimops															SgM				
Minunacovirus																			
Nyctacovirus																			
Pedacovirus																			
Colacovirus																			
Myotacovirus																			
Duvinacovirus																			
Setracovirus																			
Desmodus																			
Rhinacovirus								SgM								SgM			
Tegacovirus																P			
Minacovirus																			
Luchacovirus	SgM						SgM									SgM			
Sunacovirus																SgM	SgM		
Soracovirus																SgM			
Sarbecovirus																			
Hibecovirus																			
Nobecovirus																			
Merbecovirus																			SgM
Embecovirus																			
Igacovirus																P		P	
Brangacovirus																			
Cegacovirus																			
Buldecovirus																GM / P			
Herdecovirus																GM			
Andecovirus																GM			

Stars indicate the CoV subgenera that are implicated in recombination events throughout the analyses

Nikolaidis et al., 2021; Mol Biol Evol

Non Homologous recombination

- We built 73 protein family profiles of accessory ORFs (AOFs) using PSI-BLAST.
- With these profiles we searched for presence/absence of protein families in the various representative CoV genomes.
- Very distinct AOF architectures in the 4 genera and even in different subgenera of the same genus.
- No AOF was present in all four genera.
- Three AOFs were present in some subgenera of both α- and β-CoVs.
- Three AOFs were present in subgenera of both γ- and δ-CoVs.
- Interestingly, three of these intergenus AOFs are localized in the neighborhood of the Spike ORF.

Nikolaidis et al., 2021; Mol Biol Evol

Non Homologous recombination

We also searched in the Non-redundant NCBI database for presence of these families in other viruses or hosts (HGT).

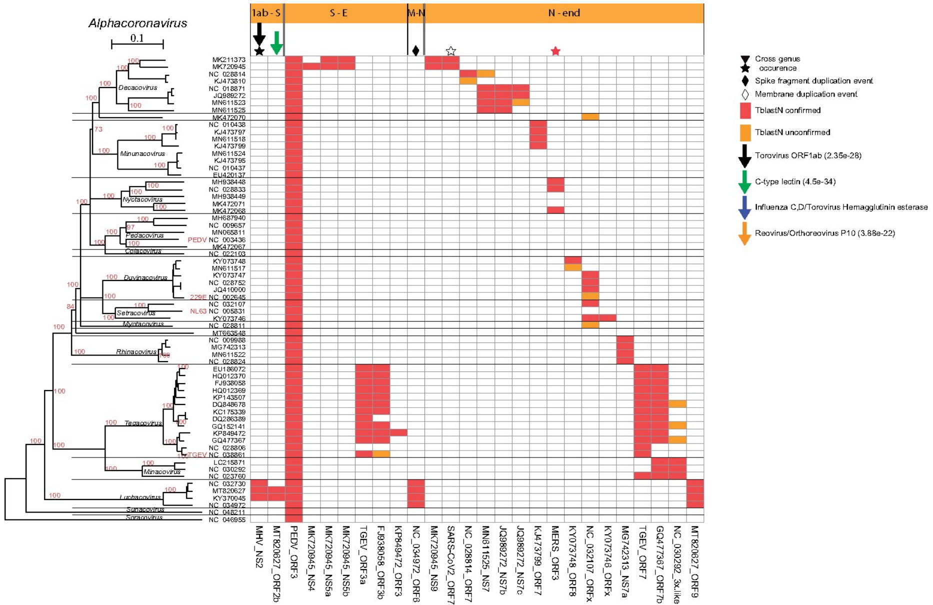
We found homologs for 7 AOFs in:

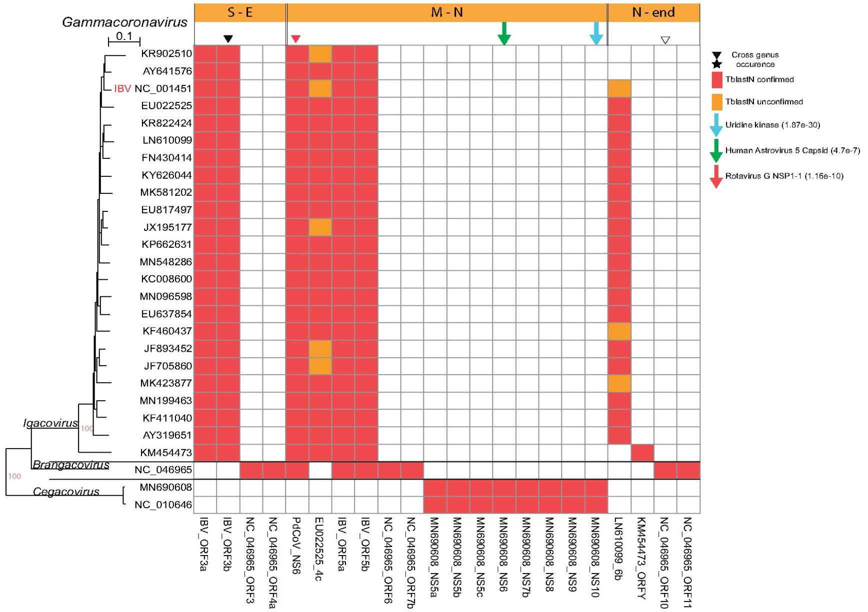
- Toroviruses
- Reoviruses
- Influenza C/D (hemagglutinin esterase)
- Avian Rotavirus-g
- Human Astrovirus 5
- Hosts (Whales, rodents)

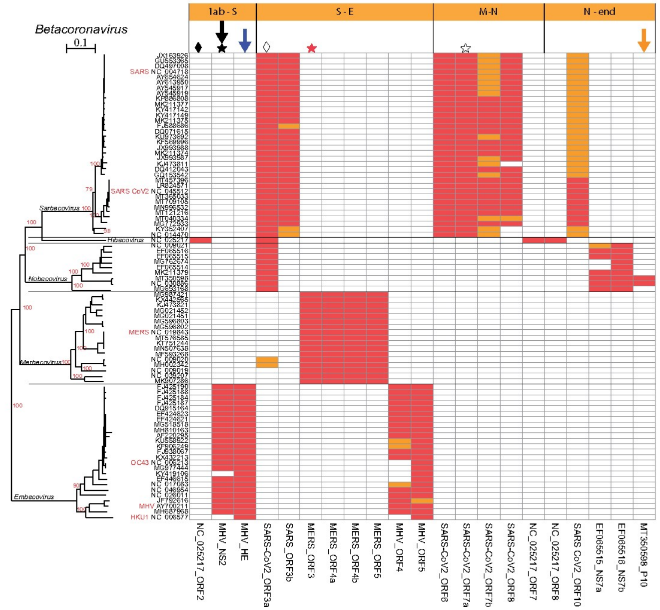
Three of these HGTs are at the borders of Spike ORF.

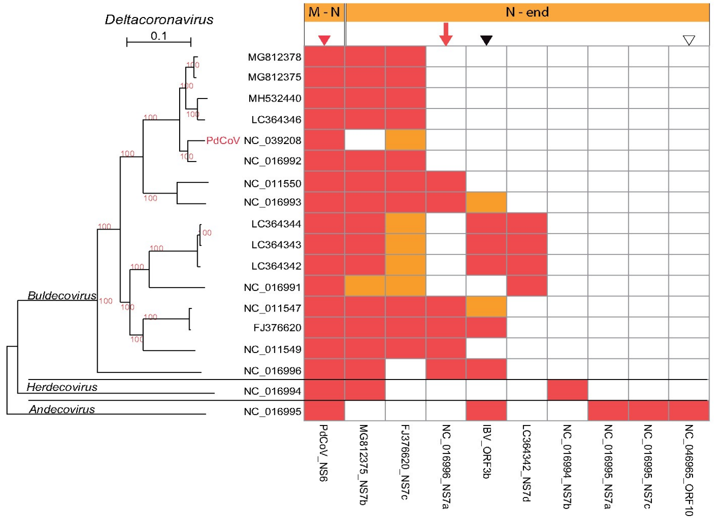
Nikolaidis et al., 2021; Mol Biol Evol

Non Homologous recombination









Main Conclusions

Most of the intertypic homologous recombinations among different subgenera occur at the Spike ORF.

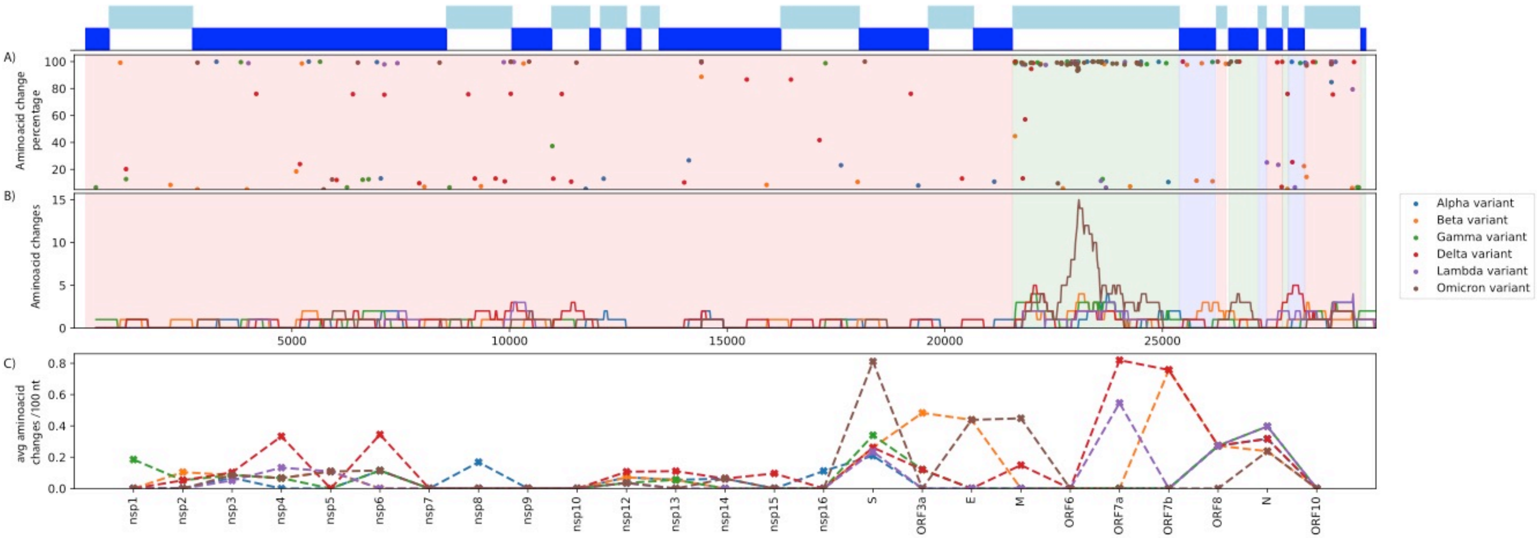
These events are not single-crossover but double crossover recombinations (cassette-like).

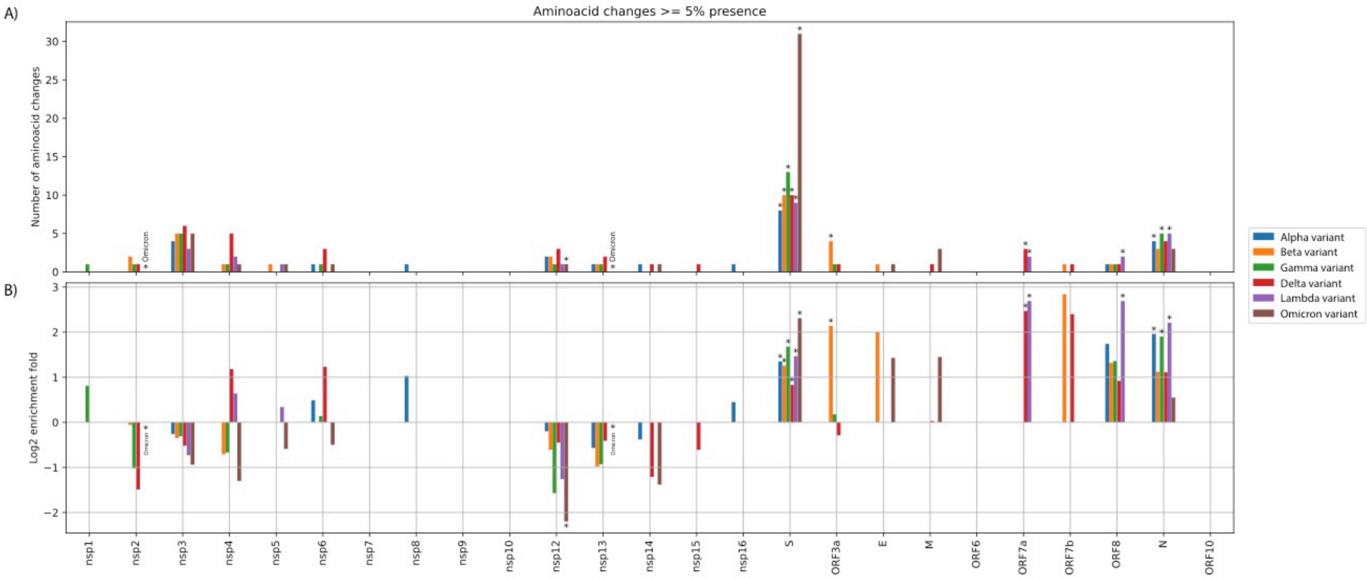
Genera and subgenera have very distinct AOF architectures. Single crossover sites would generate incompatible AOF combinations.

We did not observe intertypic events in Beta-CoVs, but this could be a matter of sampling and time.

Nikolaidis et al., 2021; Mol Biol Evol

Mutations of highly infectious VoCs

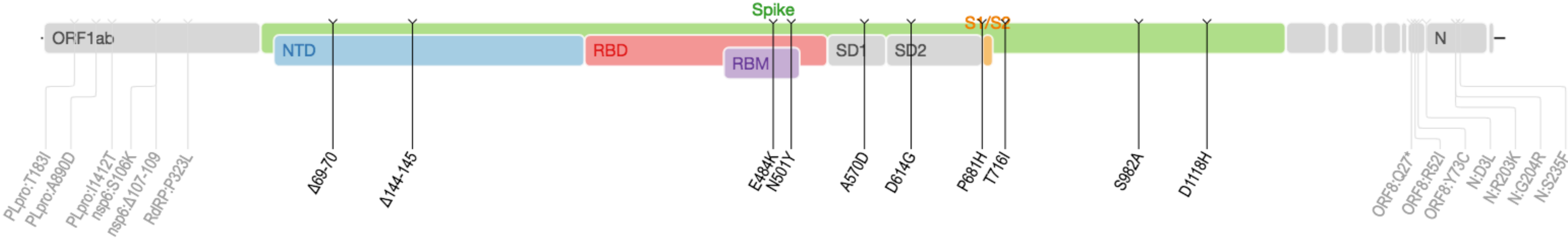




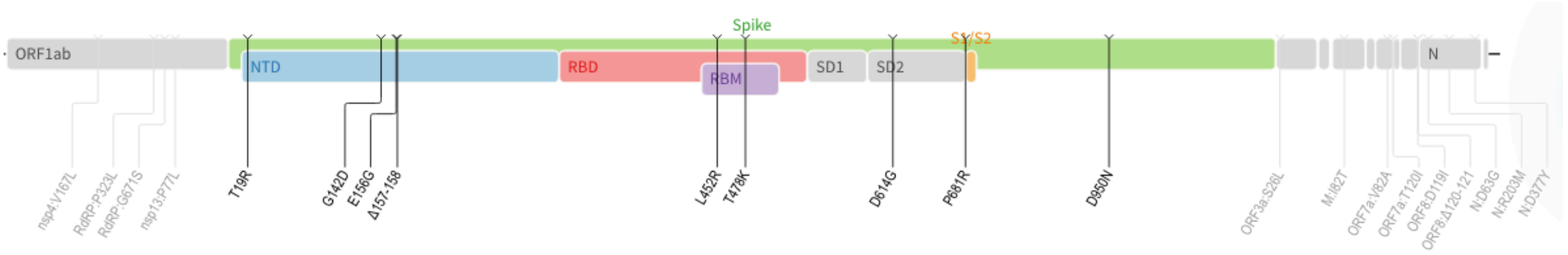
Nikolaidis et al., 2022*; Viruses*

Mutations of highly infectious VoCs

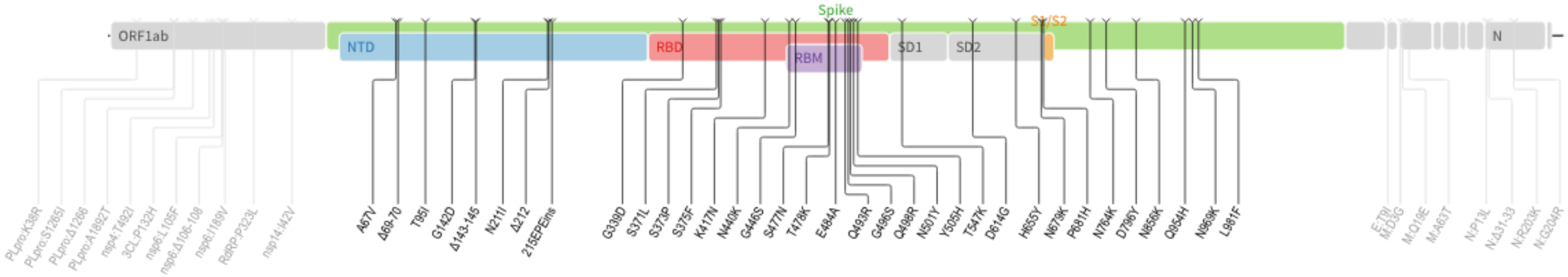
ALPHA



DELTA



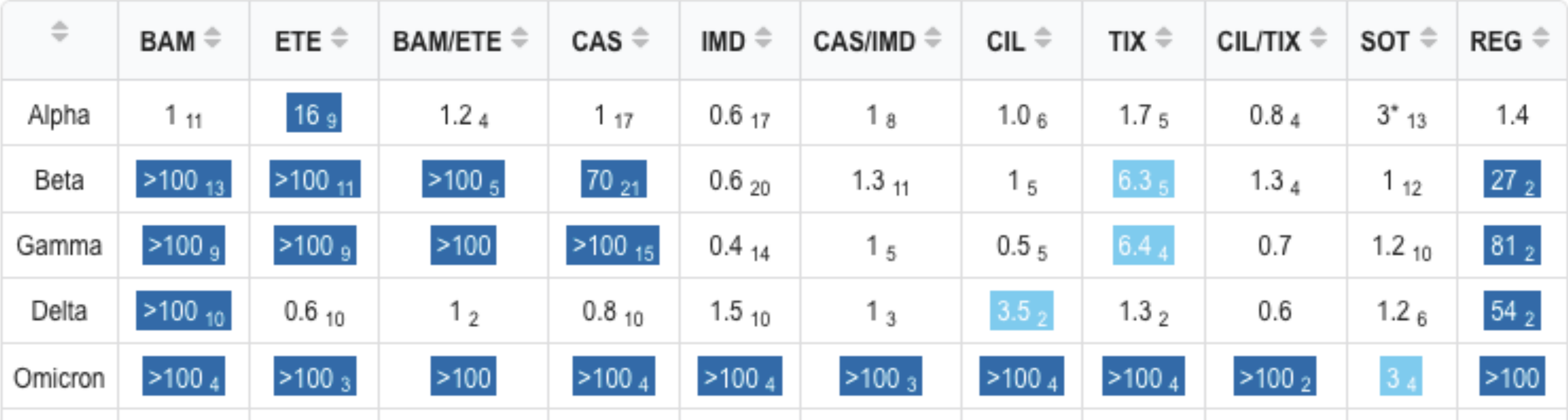
OMICRON



Coronavirus Antiviral & Resistance Database

Fold reduced neutralizing susceptibility to monoclonal antibodies under Emergency Use Authorization(EUA)

26 out of 29 existing monoclonal antibodies (mAbs) that target the receptor binding motif (RBM) cannot neutralize Omicron *in vitro*, although some mAbs that target antigenic sites outside of RBM still can.



	BAM	ETE	BAM/ETE	CAS	IMD	CAS/IMD	CIL	TIX	CIL/TIX	SOT	REG
Alpha	1 11	16 9	1.2 4	1 17	0.6 17	1 8	1.0 6	1.7 5	0.8 4	3* 13	1.4
Beta	>100 13	>100 11	>100 5	70 21	0.6 20	1.3 11	1 5	6.3 5	1.3 4	1 12	27 2
Gamma	>100 9	>100 9	>100	>100 15	0.4 14	1 5	0.5 5	6.4 4	0.7	1.2 10	81 2
Delta	>100 10	0.6 10	1 2	0.8 10	1.5 10	1 3	3.5 2	1.3 2	0.6	1.2 6	54 2
Omicron	>100 4	>100 3	>100	>100 4	>100 4	>100 3	>100 4	>100 4	>100 2	3 4	>100

Monoclonal antibody mAb) abbreviations:

BAM: Bamlanivimab/LY-CoV555,

CAS: Casirivimab/REGN10933,

IMD: Imdevimab/REGN10987,

CAS/IMD: Casirivimab+imdevimab/REGN-COV2,

ETE: Etesevimab/LY-CoV016/JS016/CB6,

CIL: Cilgavimab/COV2-2130/AZD1061,

TIX: Tixagevimab/COV2-2196/AZD8895,

TIX/CIL: Tixagevimab+Cilgavimab,

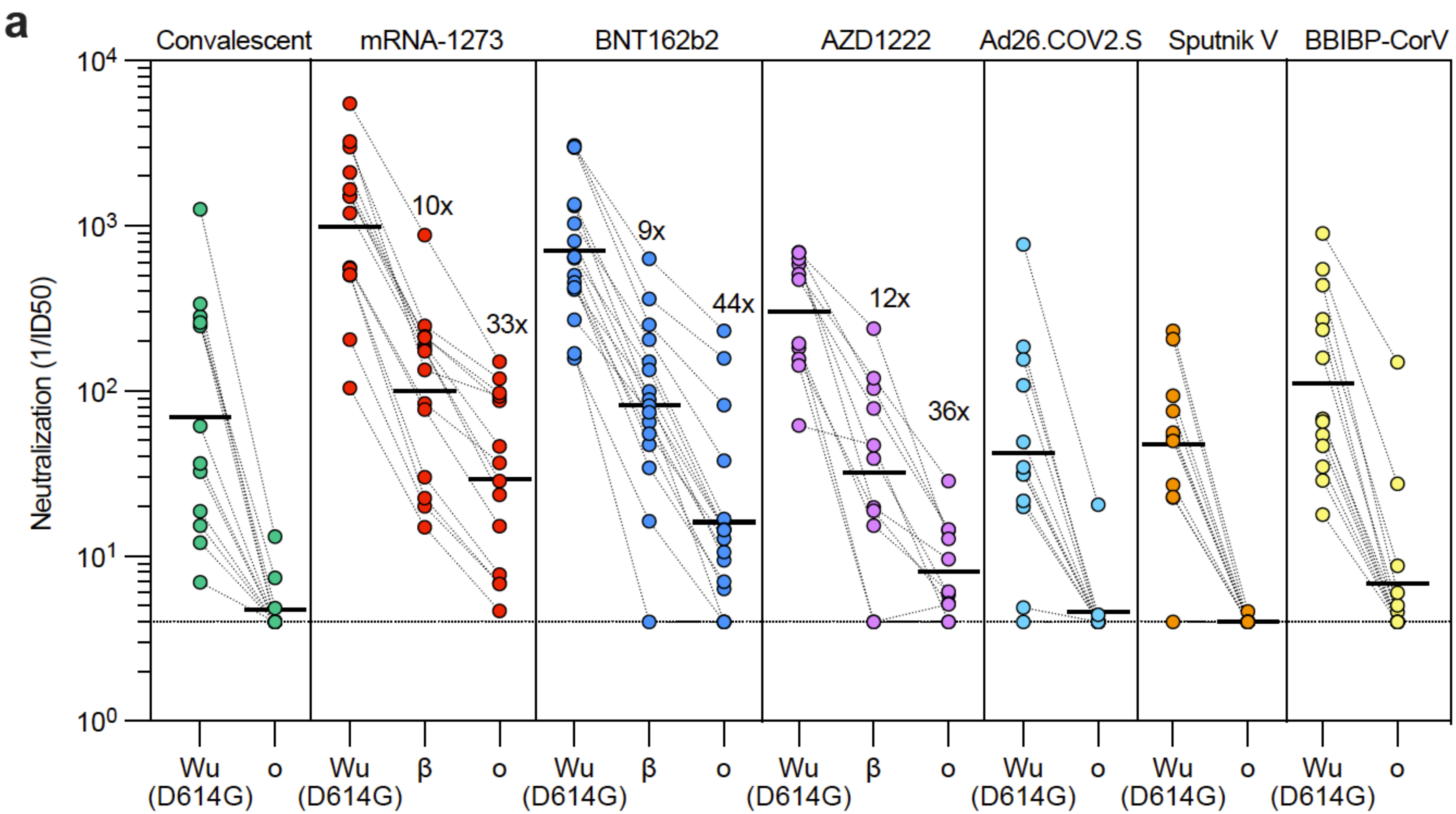
BAM/ETE: Bamlanivimab+Etesevimab,

SOT: Sotrovimab/Vir-7831/S309,

REG: Regdanvimab/CT-P59.

Coronavirus Antiviral & Resistance Database

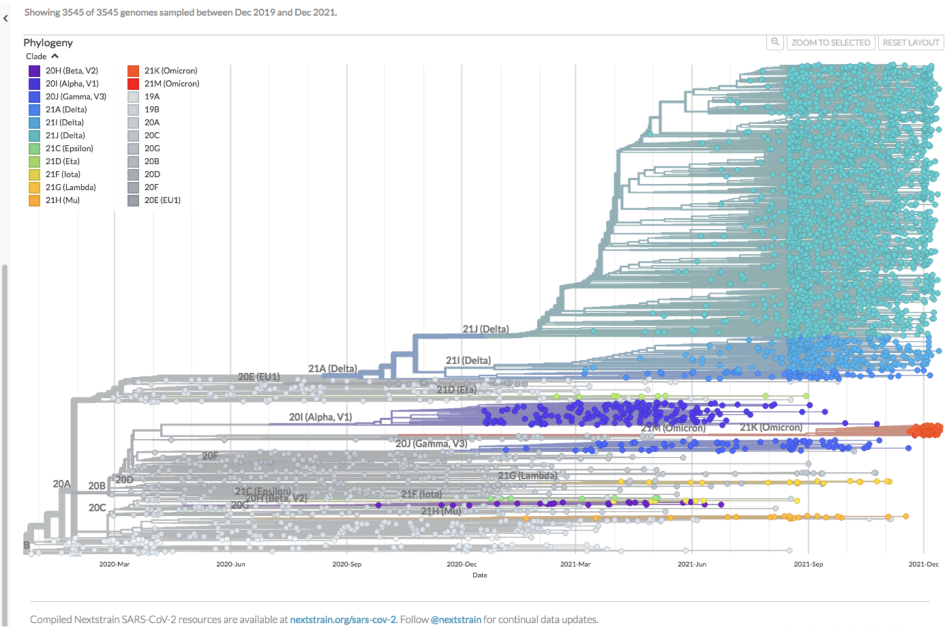
Fold reduced neutralizing susceptibility to monoclonal antibodies under Emergency Use Authorization(EUA)



Neutralization of Omicron SARS-CoV-2 VSV pseudovirus by plasma from COVID-19 convalescent and vaccinated individuals. Plasma neutralizing activity in COVID-19 convalescent or vaccinated individuals (mRNA-1273, BNT162b2, AZD1222, Ad26.COV2.S (single dose), Sputnik V and BBIBP-CorV).

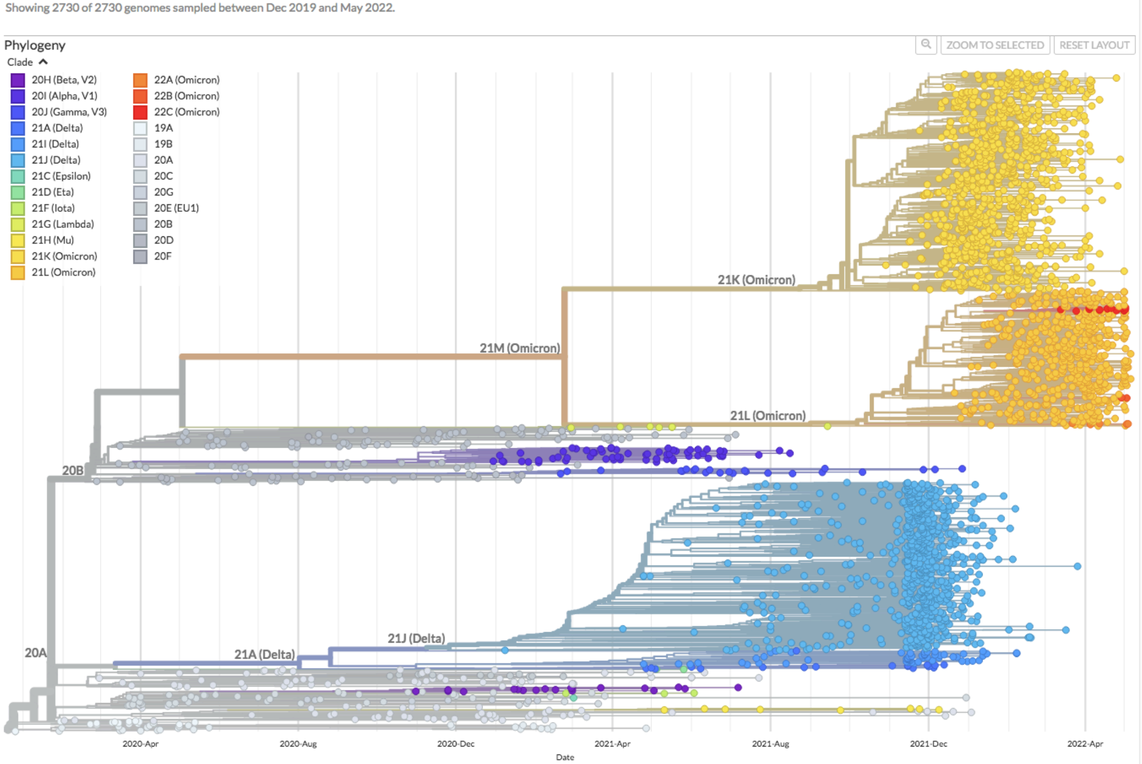
a, Pairwise neutralizing antibody titers (ID50) against Wuhan-Hu-1 (D614G), Beta and Omicron VOC. Vero E6-TMPRSS2 used as target cells. Shown one representative experiment out of 2.

Cameroni et al., 2021; Nature



Showing 3545 of 3545 genomes sampled between Dec 2019 and Dec 2021.

Compiled Nextstrain SARS-CoV-2 resources are available at nextstrain.org/sars-cov-2. Follow @nextstrain for continual data updates.



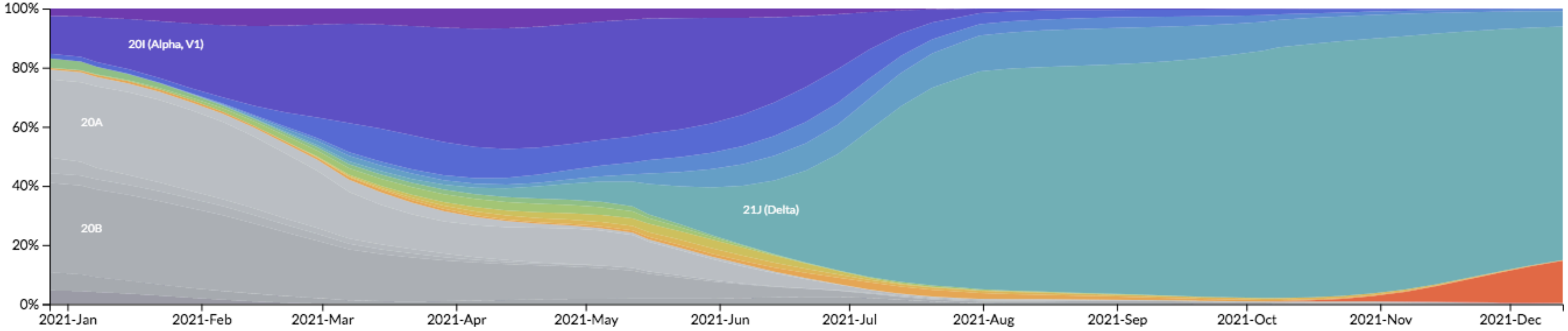
Showing 2730 of 2730 genomes sampled between Dec 2019 and May 2022.

Genomic epidemiology of novel coronavirus - Global subsampling

Built with nextstrain/ncov. Maintained by the Nextstrain team. Enabled by data from GISAID.

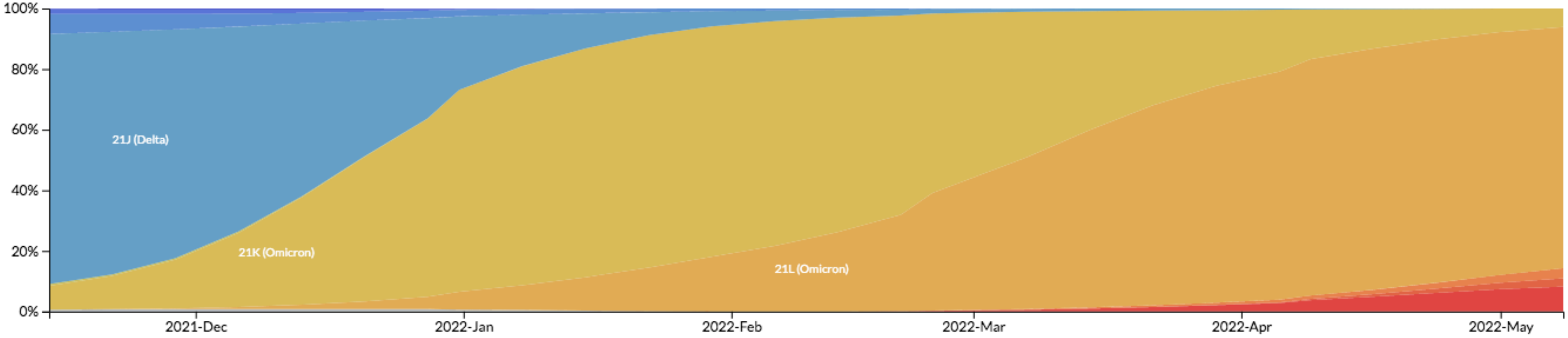
Showing 3545 of 3545 genomes sampled between Dec 2019 and Dec 2021.

Frequencies (colored by Clade)



Compiled Nextstrain SARS-CoV-2 resources are available at nextstrain.org/sars-cov-2. Follow @nextstrain for continual data updates.

Frequencies (colored by Clade)



Implications for vaccine design and development

More than twenty COVID-19 vaccines have received Emergency Use Authorization in at least one country, with efficacies in the range of 66-95%

Two major categories:

- Inactivated virus.
- Spike DNA/mRNA or Spike protein (higher efficacies).

Omicron (B.1.1.529) VoC with an unusually high number of mutations located at the Spike region raises great concern about the efficacies of current vaccines against this VoC.

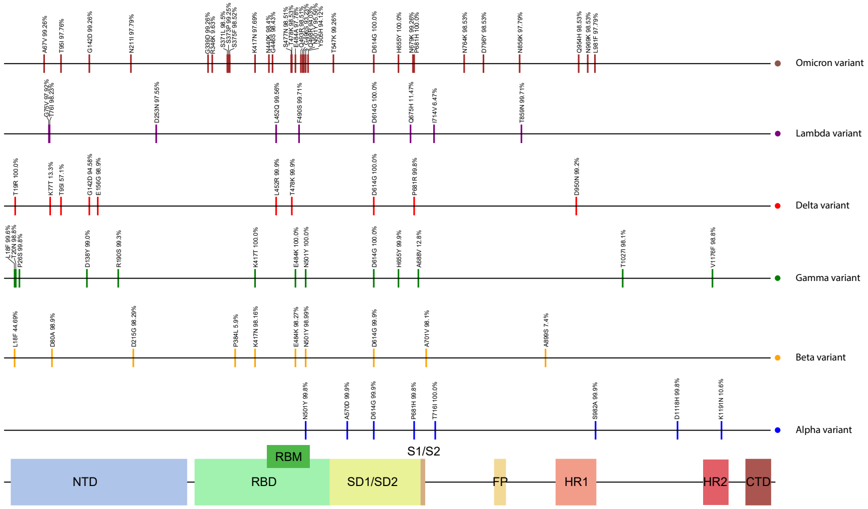
Pan-coronavirus vaccines based on Spike.

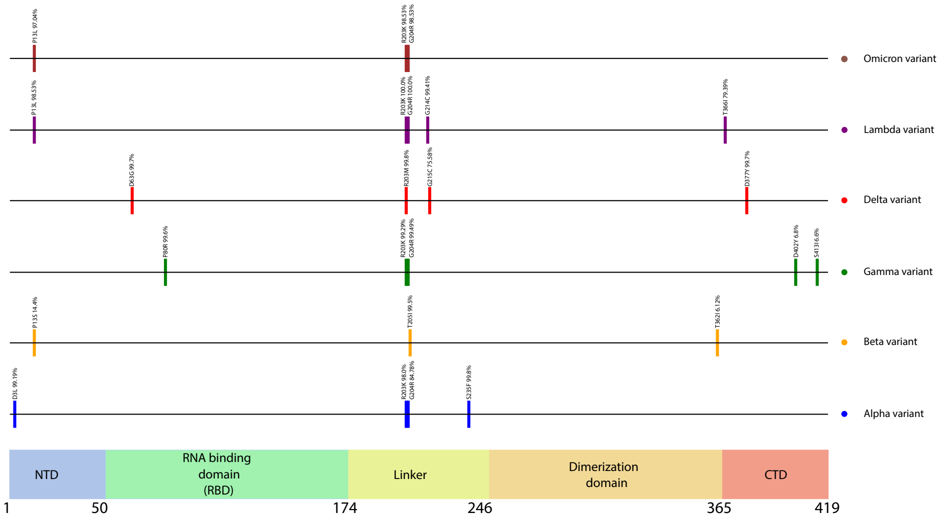
Nucleocapsid based vaccines. Three have reached Phase 2.

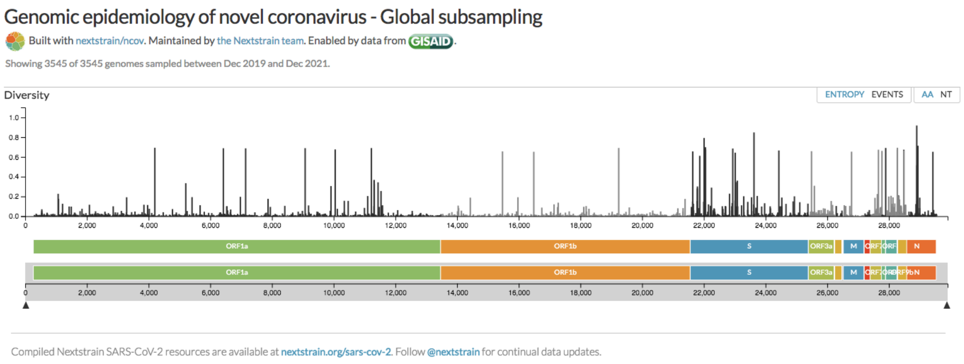
Vaccines against the early-transcribed replication transcription complex (RTC).

Implications for vaccine design and development

Mutations in Spike vs Nucleocapsid





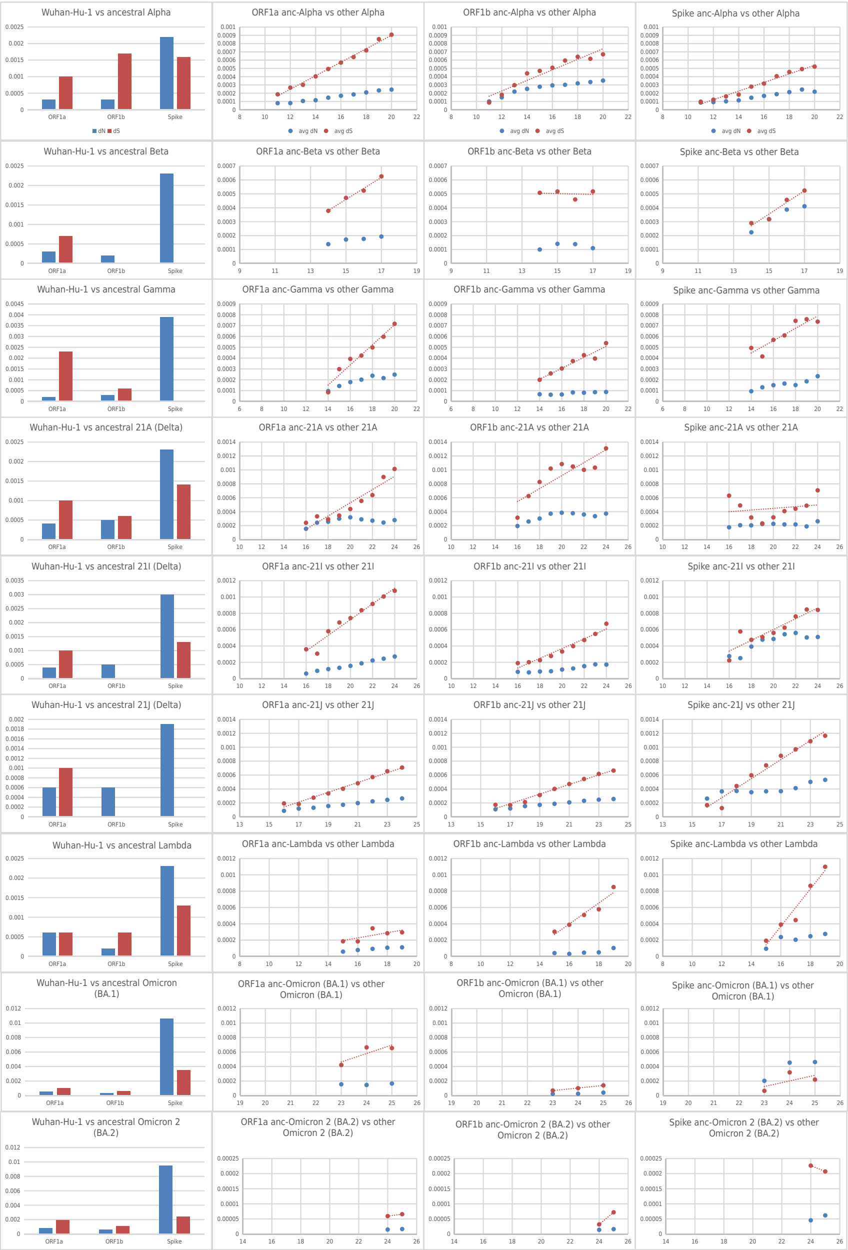


Genomic epidemiology of novel coronavirus - Global subsampling

Built with nextstrain/ncov. Maintained by the Nextstrain team. Enabled by data from GISAID.

Showing 3545 of 3545 genomes sampled between Dec 2019 and Dec 2021.

Compiled Nextstrain SARS-CoV-2 resources are available at nextstrain.org/sars-cov-2. Follow @nextstrain for continual data updates.



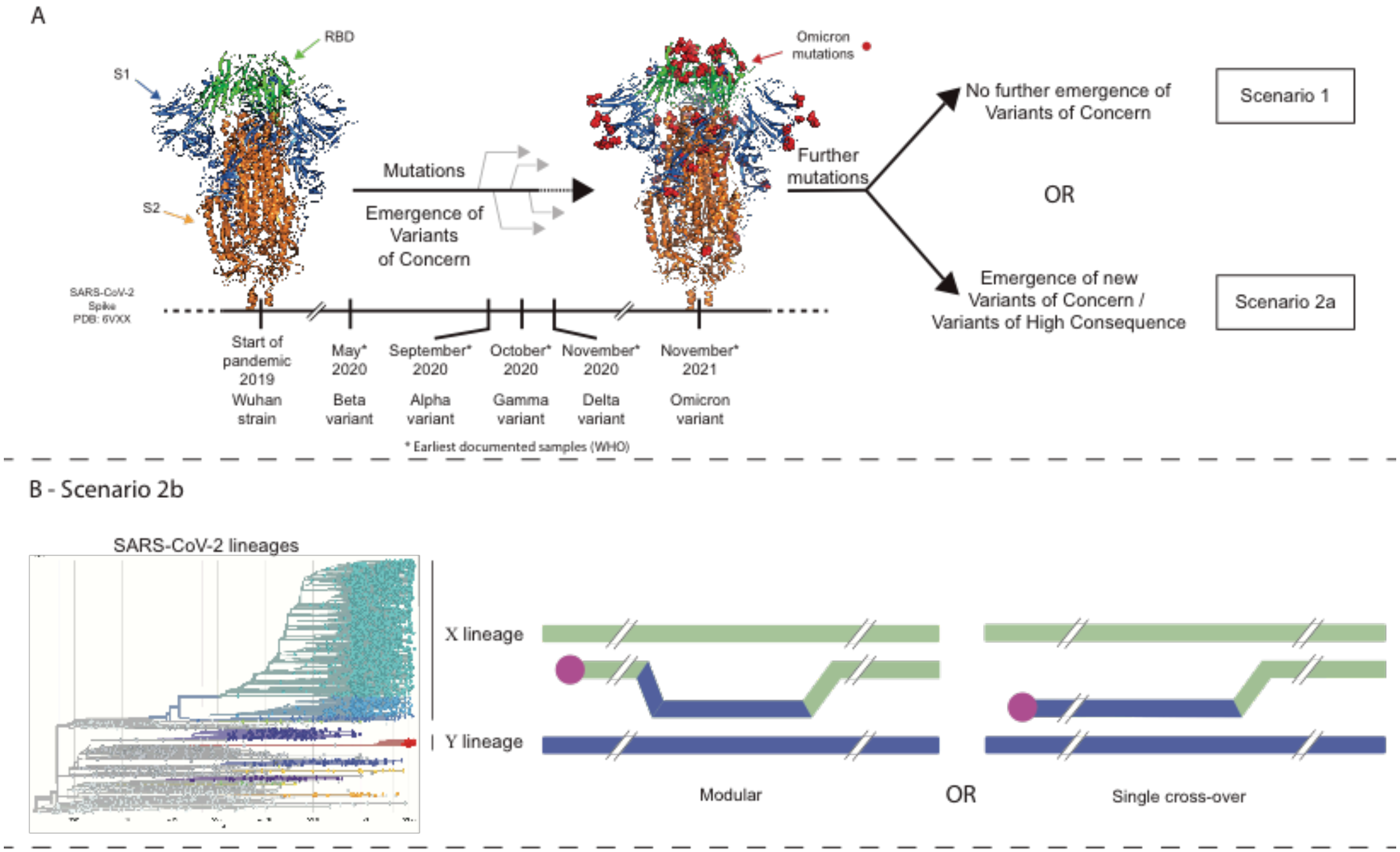
The progenitors of VoCs underwent positive selection, but once established, they underwent purifying selection

Nikolaidis et al., 2022; *Viruses*

Five scenarios for the evolution of SARS-CoV-2

Scenario 1: Structural constraints limit any further evolution of the SARS-CoV-2 Spike.
Omicron shows that this scenario is not strongly supported.

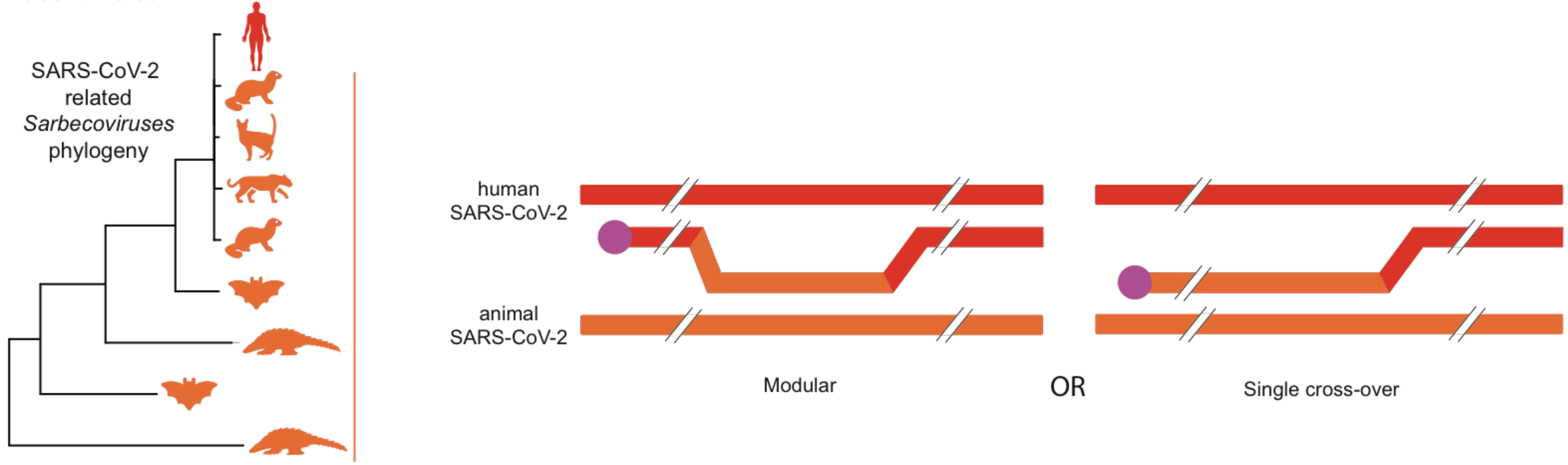
Scenario 2: Point Mutations, insertions/deletions, and/or intra-SARS-CoV-2 recombination events lead to the evolution of novel SARS-CoV-2 strains.

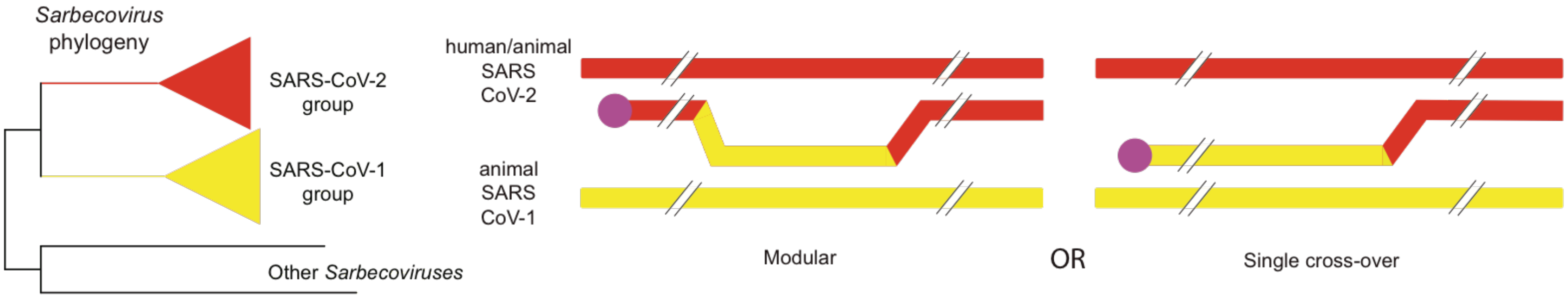


Amoutzias et al., 2021; submitted in Viruses

Five scenarios for the evolution of SARS-CoV-2

Scenario 3: Intratypic recombinations between SARS-CoV-2 and other Sarbecoviruses

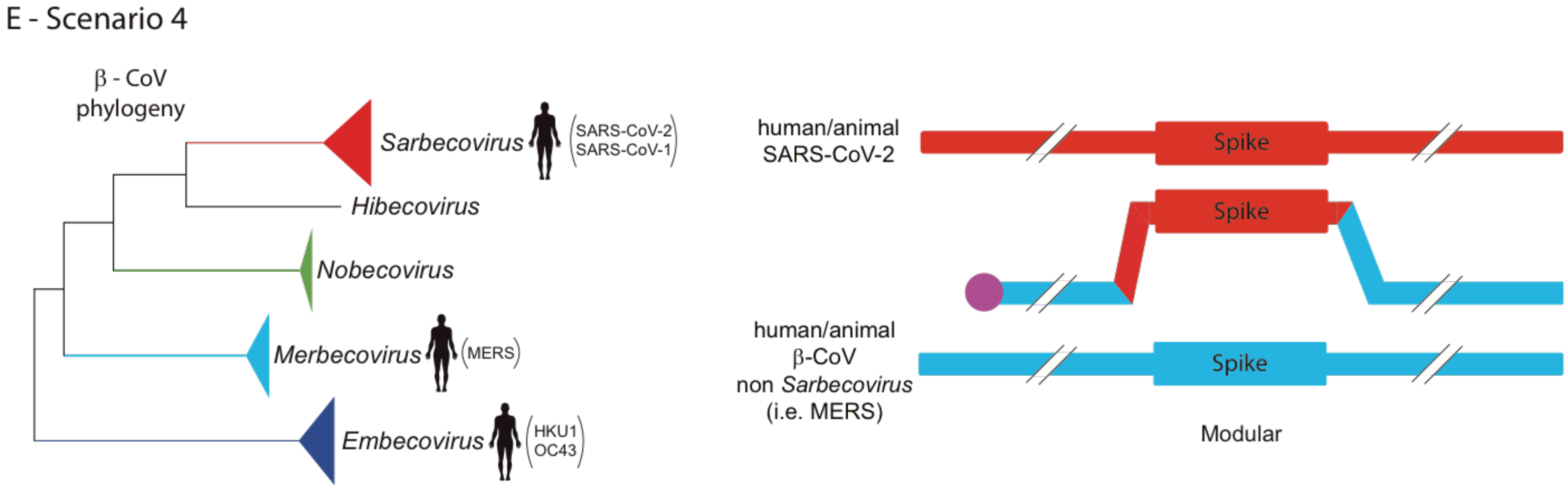

C - Scenario 3a


D - Scenario 3b

Amoutzias et al., 2021; submitted in Viruses

Five scenarios for the evolution of SARS-CoV-2

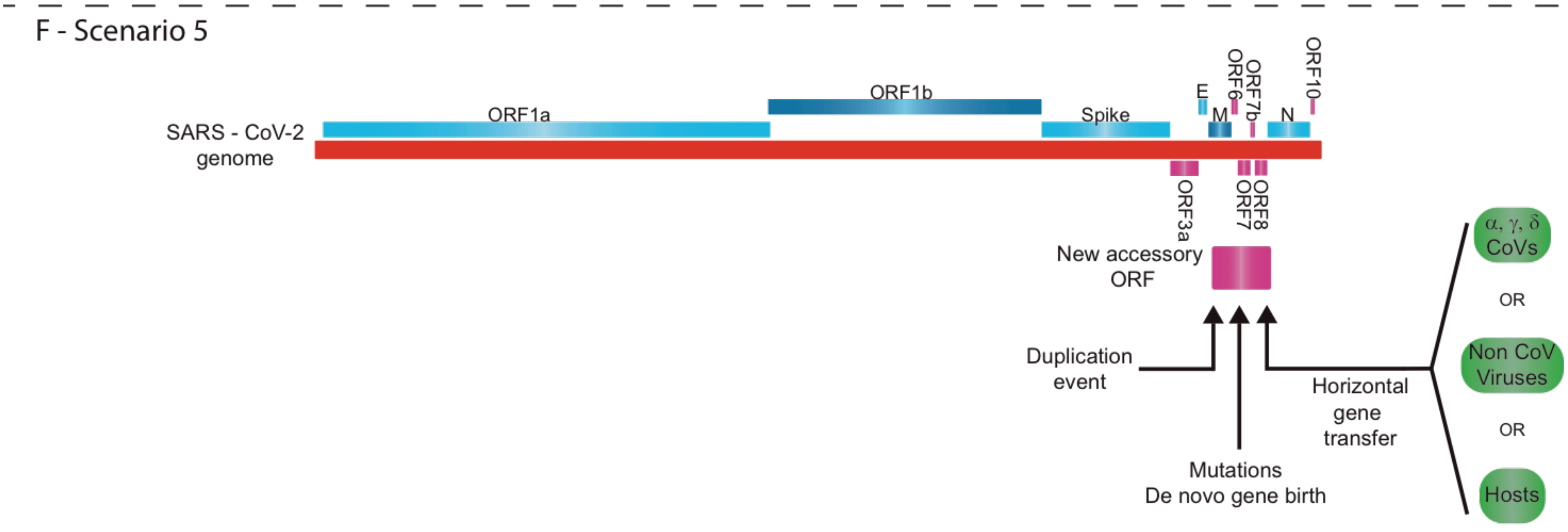
Scenario 4: Intertypic recombination between SARS-CoV-2 and viruses from other Beta-CoV subgenera



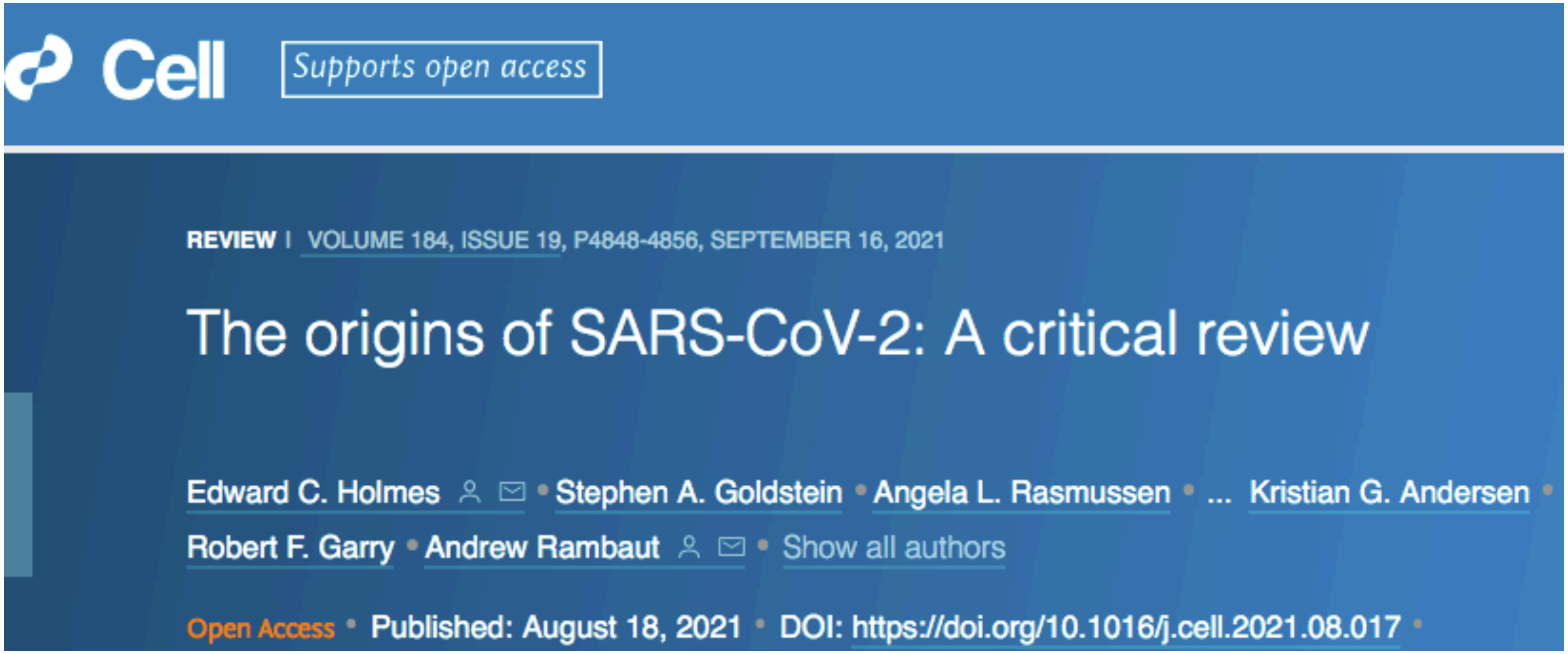
Amoutzias et al., 2021; submitted in Viruses

Five scenarios for the evolution of SARS-CoV-2

Scenario 5: Accessory ORF acquisition by non-homologous recombination of SARS-CoV-2 with other coronaviruses or even other viruses/hosts or even via *de novo* gene birth.



Amoutzias et al., 2021; submitted in Viruses



Cell

Supports open access

REVIEW | VOLUME 184, ISSUE 19, P4848-4856, SEPTEMBER 16, 2021

The origins of SARS-CoV-2: A critical review

Edward C. Holmes • Stephen A. Goldstein • Angela L. Rasmussen • ... Kristian G. Andersen • Robert F. Garry • Andrew Rambaut • Show all authors

Open Access • Published: August 18, 2021 • DOI: https://doi.org/10.1016/j.cell.2021.08.017

“We contend that although the animal reservoir for SARS-CoV-2 has not been identified and the key species may not have been tested, in contrast to other scenarios there is **substantial body of scientific evidence supporting a zoonotic origin.**

Although the possibility of a laboratory accident cannot be entirely dismissed, and may be near impossible to falsify, this conduit for emergence is highly unlikely relative to the numerous and repeated human-animal contacts that occur routinely in the wildlife trade.”

REUTERS®

October 30, 2021
1:50 PM GMT+3
Last Updated 7 months ago

World

U.S. spy agencies say origins of COVID-19 may never be known

By Mark Hosenball and Patricia Zengerle

“WASHINGTON, Oct 29 (Reuters) - U.S. intelligence agencies said on Friday they **may never be able to identify the origins of COVID-19**, as they released a new, more detailed version of their review of whether the coronavirus came from animal-to-human transmission or leaked from a lab.

The Office of the U.S. Director of National Intelligence (ODNI) said in a declassified report that a natural origin and a lab leak are both plausible hypotheses for how SARS-COV-2 first infected humans. But it said **analysts disagree** on which is more likely or whether any definitive assessment can be made at all.
”

Acknowledgments

Marios Nikolaidis – Ph.D Student, University of Thessaly, Greece

Prof. Panayotis Markoulatos, University of Thessaly, Greece

Prof. Steve Oliver, Cambridge University, UK

Prof. Yves Van de Peer, Ghent University, Belgium

Prof. Katerina Chlichlia, Democritus University of Thrace, Greece

Eleni Tryfonopoulou – Ph.D Student, Democritus University of Thrace, Greece