Emergence of a recurrent insertion in the N-terminal domain of the SARS-CoV-2 spike glycoprotein

Marco Gerdol[1*], Klevia Dishnica[2], Alejandro Giorgetti[2]

[1]University of Trieste, Department of Life Sciences, 34127 Trieste, Italy

[2]University of Verona, Department of Biotechnology, 37134 Verona, Italy

Correspondence should be addressed to: mgerdol@units.it

Abstract

Tracking the evolution of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) through genomic surveillance programs is undoubtedly one of the key priorities in the current pandemic situation. Although the genome of SARS-CoV-2 acquires mutations at a slower rate compared with other RNA viruses, evolutionary pressures derived from the widespread circulation of SARS-CoV-2 in the human population have progressively favored the global emergence, though natural selection, of several variants of concern that carry multiple non-synonymous mutations in the spike glycoprotein. These are often placed in key sites within major antibody epitopes and may therefore confer resistance to neutralizing antibodies, leading to partial immune escape, or otherwise compensate infectivity deficits associated with other non-synonymous substitutions. As previously shown by other authors, several emerging variants carry recurrent deletion regions (RDRs) that display a partial overlap with antibody epitopes located in the spike N-terminal domain (NTD). Comparatively, very little attention has been directed towards spike insertion mutations. This manuscript describes a single recurrent insertion region (RIR1) in the N-terminal domain of SARS-CoV-2 spike protein, characterized by at least 25 independent acquisitions of 2-8 additional codons between Arg214 and Asp215 in different viral lineages. Even though RIR1 is unlikely to confer antibody escape, its association with two distinct formerly widespread lineages (A.2.5 and B.1.214.2) and with known VOCs and VOIs warrants further investigation concerning its effects on spike structure and viral infectivity.

Introduction

Coronaviruses generally accumulate mutations at a much lower rate than other RNA viruses, thanks to the efficient proofreading exonuclease activity exerted by nsp14, in complex the activator protein nsp10 (Denison et al., 2011; Ma et al., 2015). As a result, the rate of molecular evolution of SARS-CoV-2 is currently estimated (as of September 1st 2021, based on GISAID data (Shu and McCauley, 2017)), to be close to 25 substitutions/year per genome, i.e. 8.36×10^{-4} substitutions/site/year, which is very close to previous estimates for human endemic coronaviruses (Ren et al., 2015). Consistently with comparative genomics data obtained from other members of the Sarbecovirus subgenus, such mutations are not evenly distributed across the genome, but they are disproportionally located in the S gene, which encodes the spike glycoprotein. It is also worth noting that the S gene undergoes frequent recombination events, likely as a result of naturally occurring co-infections in the animal viral reservoirs (Boni et al., 2020), and that these events are also theoretically possible among different SARS-CoV-2 lineages (Ignatieva et al., 2021). The encoded transmembrane protein forms a homotrimer and plays a fundamental role in the interaction between the virus and host cells, promoting viral entry though the interaction with different membrane receptors (Millet et al., 2020), which, in the case of SARS-CoV-2 and of the closely related SARS-CoV responsible of the 2002-2004 outbreak, is represented by the angiotensin converting enzyme 2 (ACE2) (Ren et al., 2008; Walls et al., 2020).

While most of these mutations have little or no phenotypic impact at all, some may significantly influence viral transmissibility and the ability of the virus to escape host immune response. The causes underpinning such phenotypic effects may either lie in an increased viral shedding, in the alteration of the binding affinity between the spike receptor binding domain (RBD) and the host ACE2 receptor, or in the modification of key antibody epitopes. The most striking example of a non-synonymous mutation which had a dramatic impact on the dynamics of the pandemics is most certainly represented by S:D614G. This mutation, which was not present in the ancestral lineage that caused the Wuhan outbreak, emerged in the very early phases of the pandemics, quickly becoming dominant worldwide (Korber et al., 2020), most likely due to an increased packing of functional spike protein into the virion (Zhang et al., 2020).

Even though the mutation rate of the SARS-CoV-2 genome remained relatively stable throughout 2020, growing evidence soon started to point out the presence of shared mutations across multiple independent lineages, suggesting ongoing convergent evolution and possible signatures of host adaptation (van Dorp et al., 2020a). While early investigations failed to identify evidence of increased transmissibility associated with such recurrent mutations (van Dorp et al., 2020b), the nearly contemporary independent emergence of three variants sharing the non-synonymous substitution N501Y in the spike protein started to raise serious concerns about the possible involvement of this mutation in increasing viral infectivity. While the functional role of N501Y still remains to be fully elucidated, structural modeling points towards a possible function in the stabilization of the spike protein in the open conformation, which may increase ACE2 binding, especially in combination with other mutations targeting the RBD (Nelson et al., 2021; Teruel et al., 2020; Zhu et al., 2021).

B.1.1.7 (the alpha variant, according to WHO labeling), one of the emerging lineages carrying N501Y, spread in southeastern England in early 2020 and quickly became the dominant variant across all Europe. Alpha was significantly more transmissible than wild-type genotypes (Davies et al., 2021), but it was not associated with significant immune escape from the neutralizing activity of convalescent or vaccinated sera (Lustig et al., 2021; G.-L. Wang et al., 2021; Z. Wang et al., 2021; Xie et al., 2021). On the other hand, some point mutations present in the spike NTD, i.e. the deletion of a codon in position 144, led to full escape from the activity of a few NTD-directed monoclonal antibodies (P. Wang et al., 2021).

Two other major lineages carrying N501Y, designated as variants of concerns (VOCs) in early 2021, i.e. B.1.351 (beta) and P.1 (gamma), were linked with major outbreaks in geographical regions with very high estimated seroprevalence, i.e. in the Eastern Cape region (South Africa) (Sykes et al., 2021) and in Manaus (Amazonas, Brazil) (Sabino et al., 2021), respectively. Both variants were characterized by a constellation of non-synonymous mutations and accelerated rates of evolution, which suggested that their selection might have occurred in immunocompromised patients with persistent viral infection (Choi et al., 2020). Among the many features shared by beta and gamma, the most remarkable one was the presence of two additional RBD mutations, i.e. E484K and K417N/K417T. The former one has been identified as a key player in antibody escape, due to its presence in a major epitope recognized by class II RBD-directed antibodies (Greaney et al., 2021; Starr et al., 2021a, 2020). Mutations of K417, located in an epitope recognized by class I antibodies, are thought to provide a minor contribution to polyclonal antibody response escape (Greaney et al., 2021) and to possibly stabilize, together with E484K and N501Y, the interaction between the RBD and the ACE2 receptor (Nelson et al., 2021).

In the spring of 2021, the lineage B.1.617.2 (delta) was internationally recognized as the fourth VOC. Like the three previously mentioned variants, delta carried several non-synonymous mutations in the S gene, including S:L452R, which is located in a major class III RBD-directed antibody epitope (Shen et al., 2021) and allows complete escape the neutralizing activity of several monoclonal antibodies (mAbs) (Starr et al., 2021b). Following its initial association with the surge of infections that occurred in India in early 2021 (Dhar et al., 2021), this variant rapidly spread worldwide and replaced alpha, which strongly suggested a higher intrinsic

transmissibility (Dagpunar, 2021), possibly due to a more efficient cleavage site between the S1 and S2 subunits (Peacock et al., 2021). At the same time, delta was also found to be endowed with significant immune escape properties, which resulted in reduced sensitivity towards the sera of convalescent and vaccinated individuals (Planas et al., 2021) and in reduced vaccine effectiveness, in particular after the first dose (Bernal et al., 2021). The trends of new infections connected with these VOCs and other emerging variants of interest (VOIs) are being closely monitored due to the possible negative impact they might have on massive vaccination campaigns (Madhi et al., 2021; Shen et al., 2021).

Among the four VOCs, three (alpha, beta and delta) carry spike deletions in the NTD. Such deletions were previously shown to specifically occur in four distinct NTD sites, named Recurrent Deletion Regions (RDR) 1, 2, 3 and 4, arising in different geographical backgrounds, in independent viral lineages. Some RDR sites display a significant overlap with known immune epitopes, suggesting that they may drive antibody escape (McCarthy et al., 2021). Comparatively, very little attention has been directed towards spike insertions, even though such events are known to have played a fundamental role in the past evolution of SARS-CoV-2 spike protein by allowing, among the other things, the acquisition of a furin-like cleavage site, which is an uncommon feature in bat coronaviruses. This short motif, which is thought to be a key pathogenicity determinant (Johnson et al., 2021), is indeed completely absent in the closely related sarbecovirus RaTG13 (Ge et al., 2016) and only partly present in the recently described RmYN02 (Zhou et al., 2020) and RacCS203 (Wacharapluesadee et al., 2021).

The present work reports the independent occurrence of at least 25 distinct insertion events at the very same NTD site, located between Val213 and Leu216, which will be hereafter referred to as Recurrent Insertion Region 1 (RIR1). The transient international spread of the A.2.5 and B.1.214.2 lineages, which both carry a three-codon insertion at RIR1 combined with multiple other non-synonymous spike mutations, together with the recent identification of RIR1 insertions in the delta lineage, suggest that more attention should be put in the near future towards the monitoring of similar codon acquisitions in known VOCs and VOIs.

Materials and methods

Sequence data analysis

All SARS-CoV-2 genome data used in this study were retrieved from GISAID (last access date September 1st, 2021) (Shu and McCauley, 2017). In detail, all available sequenced genomes belonging to the lineages A.2.5 (and the related sublineages A.2.5.1, A.2.5.2 and A.2.5.3) and B.1.214.2 were downloaded, along with associated metadata. While all GISAID entries were considered for reporting observation frequencies, only high quality genomes (i.e. those listed as “complete” and “high coverage”) associated with a sampling date were taken into account for further analysis. Genomes containing long stretches of NNNs (i.e. comprising more than 25 consecutive undetermined amino acids) were discarded. Entries belonging to the sister lineages A.2.4, B.1.214, B.1.214.1, B.1.214.3 and B.1.214.4 were also retrieved and processed as mentioned above. The reference isolate Wuhan-Hu-1 was also included for tree rooting purposes. Note that several genome sequences from Panama with sampling date anterior to November 2021 were discarded due to the unreliability of associated metadata (i.e. the sampling dates appeared to be inconsistent with the very small genetic distances with recent isolates belonging to the same lineage). Overall, the A.2.5- and B.1.214.2-focused datasets included 561 and 773 sequences, respectively.

SARS-CoV-2 genomes were analyzed with the nextstrain *augur* pipeline (https://github.com/nextstrain/augur). Briefly, nucleotide sequences were aligned with MAFFT (Katoh et al., 2002) and the resulting multiple sequence alignment was used as an input for a maximum likelihood

phylogenetic inference analysis, carried out with FastTree (Price et al., 2010) under a generalized time reversible (GTR) model of molecular evolution. The resulting tree was further refined in the *augur* environment with treetime v.0.8.1 (Sagulenko et al., 2018) using sampling date metadata, generating a time-calibrated tree. Phylogenetic trees were rooted based on the oldest available genotype, which in this case was Wuhan-Hu-1, and graphically rendered using FigTree v.1.1.4.
The global frequency of insertion and deletion mutations mapped on the SARS-CoV-2 S gene was retrieved, based on GISAID data, from https://mendel.bii.a-star.edu.sg/ (last accessed on September 1st, 2021; credit to Raphael Tze Chuen Lee). Disruptive insertion and deletion mutations (i.e. those that interrupted the open reading frame of the S gene) and insertions carrying undetermined amino acids were discarded.

System setup of coarse-grained models

The simulations on the wild-type spike protein were carried out considering the crystallographic structure deposited in PDB (accession ID: 6XR8) (Cai et al., 2020). A few missing portions were modeled with Swiss Model (Waterhouse et al., 2018) in order not to compromise the molecular dynamic properties of the protein. Homology modeling was performed to obtain the 3D structures of the spike protein of A.2.5 and B.1.214.2 with Swiss Model (Waterhouse et al., 2018), using the EPI_ISL_1502836 and EPI_ISL_1296843 GISAID entries as references. Protein structures were converted to coarse-grained Martini representations using the martinize.py script (Aguayo-Ortiz et al., 2017). The coarse-grained protein coordinates were then positioned in the center of a simulation box of size 23x23x23 nm^3.
The Martini coarse-grained force field with an Elastic Network (CG-ElNeDyn) (Aguayo-Ortiz et al., 2017) was used for running the molecular dynamics simulations through the Gromacs 2019.3 package (Van Der Spoel et al., 2005). The analyses were run using isothermal-isobaric NPT ensemble equilibrium simulations. The temperature for each group (protein, water and ions) was kept constant at 315K using V-rescale thermostat (Bussi et al., 2007) with a coupling constant of 1.0 ps. The pressure was isotropically controlled by a Parrinello-Rahman barostat (Martonák et al., 2003) at a reference of 1 bar with a coupling constant of 12.0 ps and compressibility of $3x10^{-4}$. Non-bonded interactions were used in their shifted form with electrostatic interactions shifted to zero in the range of 0-1.1 nm. A time step of 20 fs was used with neighbor lists updated every 20 steps. Periodic boundary conditions were used in the x, y and z axes. ~4µs were collected for the simulations of the wild type and mutant (i.e. A.2.5 and B.1.214.2) spike proteins, respectively. The root mean square deviation of backbone beads (RMSD), the root mean square fluctuations (RMSF) and the radius of gyration (RGYR) were calculated using the gmx rms, rmsf and gyrate modules from the Gromacs package (Van Der Spoel et al., 2005). Principal component analysis (PCA), computed with MDAnalysis, was restricted to backbone beads, as it is less perturbed by statistical noise and provides significant characterization of the essential space motions (Michaud-Agrawal et al., 2011). To visualize the direction and extent of the principal motions of the simulated systems, a porcupine plot analysis was performed using the modevectors.py script in Pymol (DeLano, 2002).

Results and Discussion

Presence of a recurrent insertion region (RIR1) in the N-terminal domain of SARS-CoV-2 spike protein
The analysis of the genomic data deposited in GISAID as of September 1st, 2021 revealed that S gene insertions (excluding those that disrupted the open reading frame) were present in just a minor fraction of all sequenced SARS-CoV-2 genomes, i.e. roughly 0.25% of the total. The impact of insertion mutations in the S gene on viral evolution has been very limited compared with deletion mutations, which are found in several

widespread VOCs and VOIs, such as alpha, beta and delta. Overall, the frequency of observation of spike deletions was approximately 700 folds higher than spike insertions. As previously reported by other authors, most deletions occur in specific sites of the N-terminal domain, named Recurrent Deletion Region (RDR) 1, 2, 3 and 4 (**Figure 1**) (McCarthy et al., 2021). This is consistent with the higher rate of mutation observed for the S1 region (which includes the NTD and RBD) in human coronaviruses compared with the more slowly evolving S2 subunit (Kistler and Bedford, 2021).

Despite their lower frequency of observation, insertions do not occur randomly in the S gene. In fact, the overwhelming majority of the insertion mutations mapped so far in SARS-CoV-2 S gene target the NTD, being in most cases (over 2,000 genomes) identified at a specific site, located between codons 213 and 216 (**Figure 1**). Due to the convergent finding of such insertions in independent viral lineages (see below), this region will be hereafter named Recurrent Insertion Region 1 (RIR1).

Even though insertions were observed in several other spike sites, RIR1 is the only one where several different insertions have independently occurred in different lineages. The other major spike insertions sites currently identifiable in SARS-CoV-2 are S:ins5F, found in a small fraction of alpha genotypes, S:ins145T, found in VOI mu (Laiton-Donato et al., 2021) and S:ins679GIAL, found in AT.1 (Klink et al., 2021) (**Figure 1**).

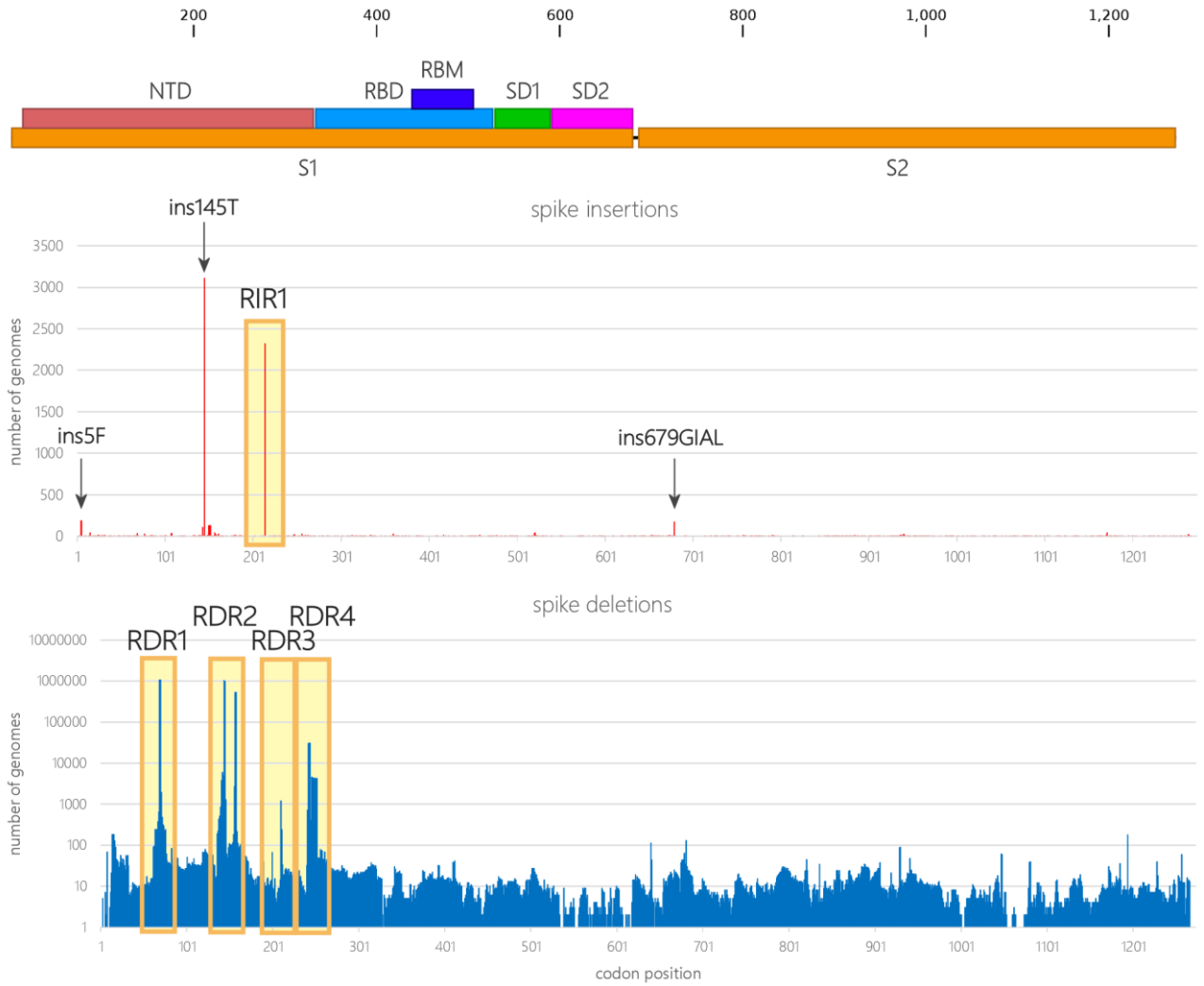


Figure 1. Schematic representation of the SARS-CoV-2 protein, with indication of the two functional S1 and S2 subunits, which are separated by a furin-like proteolytic cleavage site, the N-terminal domain (NTD), the receptor binding domain (RBD) and receptor binding motif (RBM), the SD1 and SD2 subdomains. The absolute number of observed deletion mutations along the S-gene are reported in a Log_{10} scale (https://mendel.bii.a-star.edu.sg/ was last accessed on September 1st, 2021). The position of RDR1-RDR4 from a previous study

(McCarthy et al., 2021), as well as the position of the newly identified RIR1 and of three other relevant insertions discussed in the text are reported.

RIR1 insertions were the result of at least 25 independent events that occurred in different branches of the SARS-CoV-2 phylogenetic tree, which strongly suggests convergent evolution (**Table 1**). To ensure a unique identification of each of such insertion events, these were named using Roman numerals based on their chronological identification (based on GISAID data); a lowercase letter was added to design variants of the same insertion that determined a change in the encoded amino acid sequence. Even though the length of the insertion spanned from two to eight codons (**Figure 2**), the overwhelming majority of the genomes with RIR1 insertions (99% of the total) only included three codons (**Table 1**).

While the two RIR1 insertions that led to widespread community transmission, i.e. insertions III and IV (those characterizing the A.2.5 and B.1.214.2 lineages) will be discussed in detail in the following sections, it is worth nothing that some of the other occasional events occurred in lineages that have been previously identified in VOCs or VOIs. In detail, insertion V was found in eight viral genomes belonging to the gamma lineage, sequenced in different Brazilian states and Guyana between December 2020 and April 2021, indicating the presence of community transmission in the region. As reported in a previous work (Resende et al., 2021), these genomes belong to a monophyletic P.1-like clade that appears to be basal to P.1.

The highly transmissible alpha lineage, which rapidly became dominant in Europe and quickly spread worldwide in early 2021 (Volz et al., 2021), before the rise of delta, was associated with at least five independent insertions at RIR1 (**Table 1**, **Figure 2**). These included (i) insertion XI, documented in three patients in Scotland in February 2021; (ii) insertion XII, identified in two patients in Czech Republic in February 2021; (iii) insertion XIII, identified in a single patient in Illinois, in March 2021; (iv) insertion XV, found in a single case in Greece, in late March 2021; (v) insertion XVIII, found in a single case in Florida, in May 2021. A single case of RIR1 insertion (XXV) was also recently recorded in the VOI eta (B.1.525) in the United Kingdom (Ozer et al., 2021). Three of the most recently identified insertions at RIR1 were associated with delta (**Table 1**, **Figure 2**). While insertion XXII was linked with a single case detected in Italy in July 2021, both insertion XXIII and XXIV gave rise to small clusters of community spread in England, which may be still ongoing as of September 1st, 2021.

Albeit not directly linked with variants presently designated as VOCs or VOIs, other RIR1 insertions were associated with the presence of immunologically relevant spike mutations. This is the case of insertion VII, reported in two patients in California in January 2021, which occurred the B.1.429 (epsilon) lineage, long considered to be a VOI by WHO (Zhang et al., 2021) (**Figure 2**). Moreover, insertion IX occurred in an unclassified B.1 sublineage, associated with a small cluster of infections in New York and New Jersey, characterized by the presence of E484K and by the deletions Δ69/Δ70 (found in RDR1) and Δ144 (found in RDR2), shared by several VOCs and VOIs. Due to the presence of this constellation of spike non-synonymous mutations, this variant has been placed among those under monitoring by Public Health England (Public Health England, 2021).

Taking into account the limited efforts carried out by several countries in genomic surveillance, the insertions reported in **Table 1** and **Figure 2** may represent just a fraction of those that emerged at RIR1 during the course of the pandemics. While most insertions at RIR1 were in-frame, occurring at phase 0 between codons 214 and 215, many others were out-of-frame, occurring either at phase I (i.e. between the first and the second nucleotide of a codon) or phase II (i.e. between the second and the third nucleotide of codon) (**Figure 2**). This particular placement of the insertion may determine a non-synonymous mutation of the residues flanking RIR1 at the N- and C-terminal sides (**Table 1**).

Designation	Insertion	insertion type	Lineage	GISAID entries	Other spike mutations	Earliest detection	Latest detection*
I	AKKN	in-frame	B	1	none	Mar 5th, 2020	Mar 5th, 2020
II	KLGP	in-frame	B.1.177	1	E154K, A222V, D614G	Nov 13th, 2020	Nov 13th, 2020
III	AAG	out-of-frame	A.2.5	1171	L141del, G142del, V143del, D215Y, L452R, D614G	Nov 20th, 2020	Aug 12th, 2021
IIIa	AVG	out-of-frame	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	Feb 18th, 2021	Feb 18th, 2021
IIIb	ASG	out-of-frame	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	Jan 21st, 2021	Jan 21st, 2021
IIIc	AAV	out-of-frame	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	May 17th, 2021	May 17th, 2021
IIId	DGR	out-of-frame	A.2.5	1	L141del, G142del, V143del, D215L, L452R, D614G	Jan 19th, 2021	Jan 19th, 2021
IIIe	AAGAAG	out-of-frame	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	May 6th, 2021	May 6th, 2021
IV	TDR	in-frame	B.1.214.2	1126	Q414K, N450K, D614G, T716I	Nov 22nd, 2020	June 28th, 2021
IVa	IDR	in-frame	B.1.214.2	1	Q414K, N450K, D614G, T716I	Apr 3rd, 2021	Apr 3rd, 2021
IVb	QID	in-frame	B.1.214.2	3	Q414K, N450K, D614G, T716I	Apr 12th, 2021	May 1st, 2021
V	ANRN	in-frame	P.1	8	L18F, P26S, D138Y, K417T, E484K, N501Y, D614G, D1139H, V1176F	Dec 23rd, 2020	Apr 5th, 2021
VI	KRI	in-frame	B	4	V367F, E990A	Dec 28th, 2020	Mar 15th, 2021
VII	AQER	in-frame	B.1.429	2	S13I, P26S, S98F, W152C, L452R, D614G, T1027I	Jan 15th, 2021	Jan 18th, 2021
VIII	DLA	in-frame	B.1.2	5	D614G	Jan 17th, 2021	Feb 2nd, 2021
IX	QAS	in-frame	B.1	19	H69del, V70del, Y144del, M153T, T478K, E484K, D614G, T859N, D936Y	Jan 19th, 2021	Jul 27th, 2021
X	DRS	out-of-frame	B.1	1	D215N, V382L, D614G, M1237I	Feb 1st, 2021	Feb 1st, 2021
XI	KFH	in-frame	B.1.1.7	3	H69del, V70del, Y144del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Feb 12th, 2021	Feb 22nd, 2021
XII	KAFKQ	in-frame	B.1.1.7	2	H69del, V70del, Y144del, A262S, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Feb 25th, 2021	Feb 25th, 2021
XIII	ER	out-of-frame	B.1.1.7	1	H69del, V70del, Y144del, D215G, N501Y, A570D, D614G, P681H, I712V, T716I, S982A, D1118H	Mar 11th, 2021	Mar 11th, 2021
XIV	NWAHW	in-frame	B.1.547	1	T19I, T95I, H69del, V70del, D614G, E484A, A879T, T1027I	Mar 22nd, 2021	Mar 22nd, 2021
XV	APR	in-frame	B.1.1.7	1	H69del, V70del, Y144del, A262S, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Mar 31st, 2021	Mar 31st, 2021
XVI	KLRS	out-of-frame	A.28	6	T76I, D215S, N501T, F592S, H655Y	Apr 23rd, 2021	Jun 22nd, 2021
XVII	KGE	in-frame	B.1.1.519	1	T732A, T478K, D614G, P681H	Apr 24th, 2021	Apr 24th, 2021
XVIII	QMR	out-of-frame**	B.1.1.7	1	L5F, S13I, H69del, V70del, Y144del, D215R, E484K, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	May 25th, 2021	May 25th, 2021
XIX	KLLIRGD	in-frame	B.1	2	Q14del, C15del, V16del, N17del, L18del, W64R, T95I, C136Y, N137del, L141del, G142del, V143del, W152R, I210del, G252V, T415A, N440T, E484K, D614G, H655Y, P681H, T859N, Q1011H, G1219C	Jun 4th, 2021	Jun 4th, 2021
XX	KRLR	out-of-frame***	B.1.1	1	R214Q, D614G, E484D	Jun 6th, 2021	Jun 6th, 2021
XXI	QCEE	in-frame	B.1.247	1	P209S, A222V, T572I	Jul 15th, 2021	Jul 15th, 2021
XXII	ELCD	in-frame	AY.12	1	T19R, T95I, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Jul 17th, 2022	Jul 17th, 2022
XXIII	VPWI	in-frame	AY.4	9	T19R, T95I, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R,V622F, D950N	Aug 4th, 2020	Aug 21st, 2021
XXIV	HSG	in-frame	AY.4	5	T19R, T95I, D138H, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Aug 4th, 2021	Aug 21st, 2021
XXV	SYCTKSKC	in-frame	B.1.525	1	A67V, H69del, V70del, Y144del, E484K, D614G, A653V, N679del, Q677H, F888L	Aug 5th, 2021	Aug 5th, 2021

*as of September 1st, 2021; **insertion at position 215; ***insertion at codon 213

Table 1. Summary of the 25 independent RIR1 insertions found in the SARS-CoV-2 genome, ordered by the earliest date of detection.

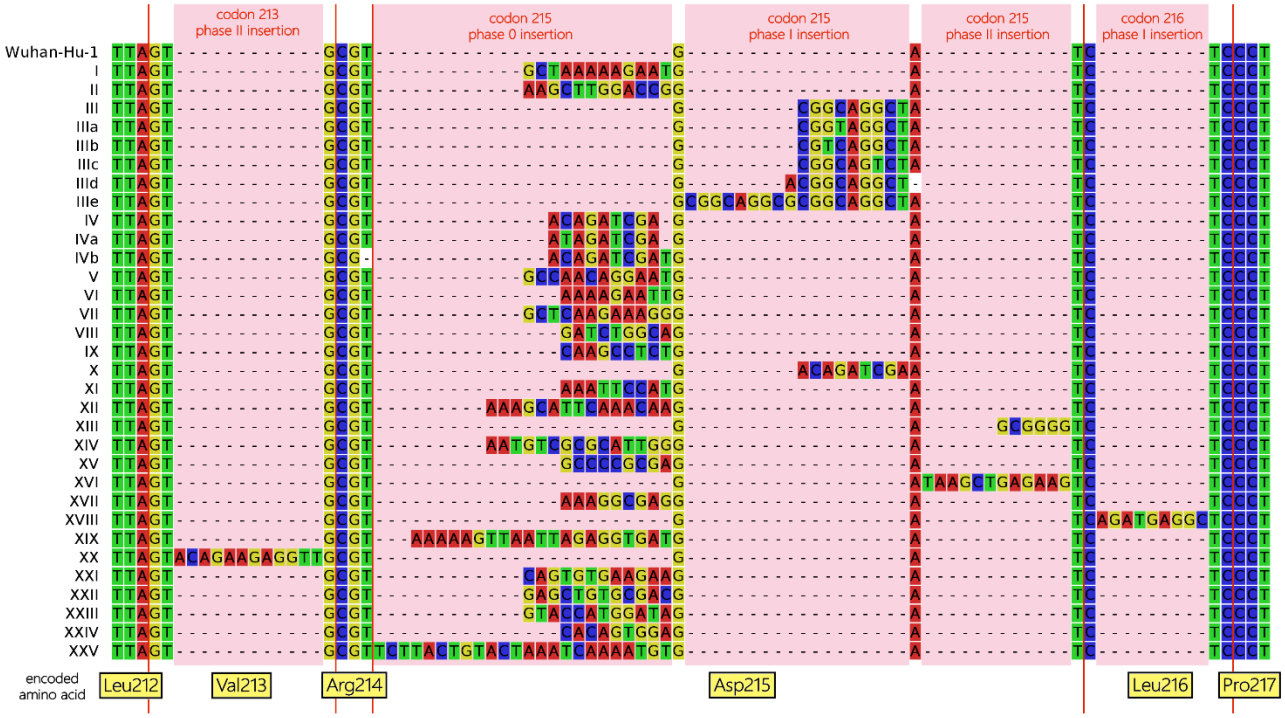


Figure 2. Multiple sequence alignment of the nucleotide sequences of the SARS-CoV-2 S gene of the viral lineages characterized by an insertion at RIR1, compared with the reference sequence Wuhan Hu-1. The multiple sequence alignment only displays a small portion of the S gene and of the encoded spike protein, zoomed-in and centered on RIR1. Red vertical bars indicate codon boundaries, with the encoded amino acids (in the Wuhan Hu-1 reference sequence) indicated below. See **Table 1** for the amino acid sequences encoded by the insertions.

Characterization of the lineage-defining mutations in A.2.5 and B.1.214

The only two lineages with insertions at RIR1 with solid evidence of widespread community transmission are A.2.5 and B.1.214.2. The inserted amino acid sequence found in A.2.5 is AAG, as the result of the phase I out-of-frame insertion of the nucleotide sequence CGTCAGGCTA, which determines the non-synonymous substitution of Asp215 to Tyr (as a result of a GAT->TAT codon replacement). In a few cases the inserted sequence was duplicated (IIIe), contained a non-synonymous mutation (IIIa, IIIb and IIIc), or was subject to more complex rearrangements (IIId) (**Table 1**, **Figure 2**).

Besides the insertion at RIR1, A.2.5 also displays the deletion of three codons (Δ141-143) in RDR2 (**Table 1**), sometimes extending to codon 144. This region has been previously implicated in antibody escape (McCarthy et al., 2021) and shows deletions in some relevant VOCs and VOIs, including alpha and eta. In particular, Δ144 appears to largely explain the resistance towards several NTD-directed mAbs displayed by alpha *in vitro* (Wang et al., 2021). Moreover, the insertion at RIR1 is also combined with L452R, a key mutation that confers resistance towards class III RDB-directed antibodies (Greaney et al., 2021), including LY-CoV555, the basis for the formulation of the commercial mAb bamlanivimab developed by Eli Lilly (Starr et al., 2021). Among the VOCc and VOIs, L452R is also found in delta and kappa. Like the overwhelming majority of the variants circulating in 2021, A.2.5 is also characterized by the presence of the prevalent mutation D614G. The summary of the mutations found in A.2.5 is reported in **Figure 3**.

The typical insertion found in the lineage B.1.214.2 is TDR, which corresponds to an in-frame insertion of the ACAGATCGA nucleotide sequence. This is combined with four non-synonymous mutations (**Figure 3**): besides D614G, Q414K and N450K are found in the RBD and T716I, which is shared by alpha, is found in the S2 region. As far as the two RBD mutations are concerned, Q414K is located close to K417, a key site mutated in the

VOCs gamma (K417T) and beta (K417N), as well as in the delta-related lineages AY.1 and AY.2 (K417N). While it is unclear whether this residue is also part of an epitope recognized by class I antibodies, whose neutralizing activity is escaped by the two aforementioned VOCs (Greaney et al., 2021; P. Wang et al., 2021), computational studies have suggested that Q414K may moderately increase RBD stability (Teng et al., 2021). On the other hand, N450 falls within a class III neutralizing antibody epitope (Yin et al., 2021), and the N450K mutation in particular has been previously associated with immune escape both towards a few mAbs and towards convalescent sera (Liu et al., 2021). Moreover, both deep mutational scanning experiments and computational simulations suggest that N450K may determine a mild increase in ACE2 binding affinity (Alouane et al., 2020; Starr et al., 2020). Currently, neither of the two B.1.214.2-associated RBD mutations are associated with other widespread SARS-CoV-2 VOCs or VOIs. It is also worth noting that B.1.214.2 displays a large deletion (10 codons) in the N-terminal region of ORF3a, which is not shared by other prominent SARS-CoV-2 lineages (**Figure 3**).

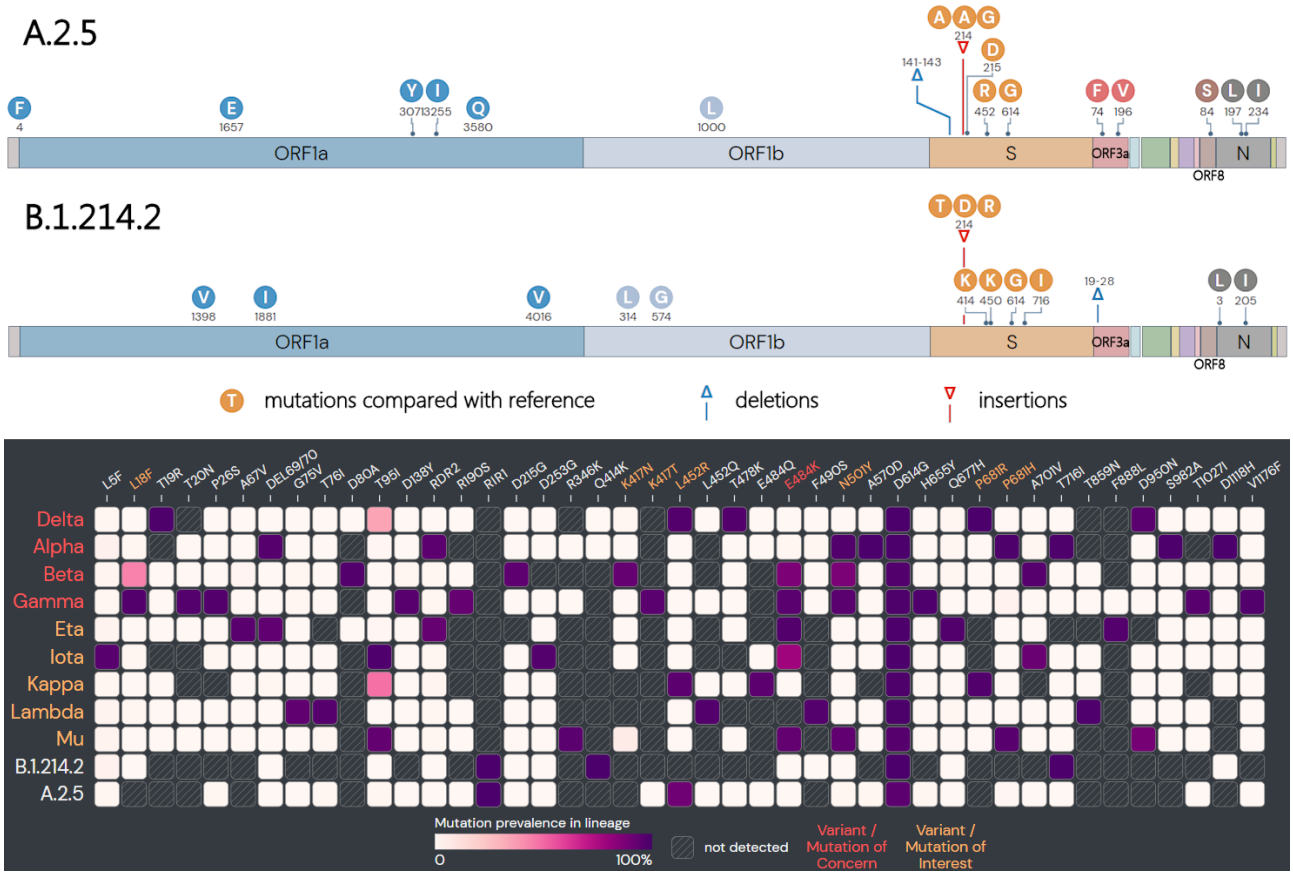


Figure 3. Upper panel: Key mutations associated with the A.2.5 and B.1.214.2 lineages. Genes associated with mutations (compared with the reference strain Wuhan-Hu-1) are indicated. Lower panel: comparison between the mutations observed in the S gene of B.1.214.2, A.2.5 and the variants currently identified as VOCs and VOIs by WHO. Modified from https://outbreak.info/.

Context for the emergence of A.2.5

A.2.5 stems from A.2.4, the dominant lineage in the Panama pandemics during the first half of 2020 (Franco et al., 2020). The first documented cases can be traced back to late November 2020, all within a 100 km^2 area around the capital city Panamá. However, the precise timing of the emergence of A.2.5, along with the acquisition of insertion III at RIR1 and of the other associated mutations described in the previous section, is presently unclear due to the insufficient molecular surveillance carried out in Central America. To date, less than 900 genomes out of nearly 460K covid-19 cases reported in this country have been sequenced i.e. less than 0.2% of the total, far below of the threshold that would be sufficient to track emerging variants (Vavrek

et al., 2021). Nevertheless, A.2.5 underwent an evident expansion in Panama between December 2020 and February 2021, as revealed by the increase in estimated prevalence from ~60 to ~95%.

Not surprisingly, the remarkable spread of SARS-CoV-2 in the country (as of September 2021, the incidence of infections surpassed 100,000 cases per million inhabitants) was connected with a significant number of exported cases, which have sometimes led to clusters of infection abroad. The A.2.5 lineage likely spread very early also in the neighboring countries, with evidence of multiple introductions. While early investigations failed to identify A.2.5 in Costa Rica in August 2020 (Molina-Mora et al., 2020), the prevalence of this lineage may have reached 30% between March and June 2021, with evidence of multiple introductions and the establishment of large clusters of community transmission, such as the one characterizing the A.2.5.1 sublineage (**Figure 4**).

The first evidence of the detection of A.2.5 outside from Central America dates to December 1st 2020, in Ecuador, where the subsequent acquisition of the spike mutation S477N, known to strengthen the binding with the ACE2 receptor (Singh et al., 2021), defined the A.2.5.3 sublineage (**Figure 4**). Reports in other Latin American countries remain sporadic, but it is worth noting that A.2.5 genomes have been so far sequenced in Argentina, Suriname, Guatemala, Honduras, Dominican Republic, Sint Marteen, Cayman Islands, Chile, Colombia, Mexico and Paraguay. The earliest cases exported in other continents have been reported with similar timing in UAE (December 27th, 2020), Philippines (December 30th, 2020) and Australia (January 2021), which is consistent with the period with the highest incidence of covid-19 infections documented in Panama. Cases linked with A.2.5 in Europe were first identified in Luxembourg, Portugal and Germany, and later in the United Kingdom, Czech Republic, Netherlands and Spain, but in most cases these did not lead to significant community transmission, with the exception of the cluster of cases linked with the A.2.5.2 sublineage recorded in Campania (central Italy) in February-March 2021. Similarly, imported cases have most certainly led to local cluster of infections in different areas of the United States, with evidence of multiple independent introductions both in New York (**Figure 4**) and Florida.

A.2.5 belongs to one of the very few surviving children lineages of the ancestral lineage A, which, after several months of limited global spread, has led to a few major clusters of infections in 2021. Such events, that occurred for example in Uganda (A.23.1) (Bugembe et al., 2021) and, most likely, also in Tanzania (A.VOI.V2) (Bugembe et al., 2021), are characterized by the acquisition of several spike mutations, often shared by other VOCs and VOIs. While the global frequency of observation of A.2.5 and related sublineages decreased after the global spread of delta, as of September 1st, 2021 this lineage is likely still circulating in Sothern, Central and Northern America (**Figure 4**).

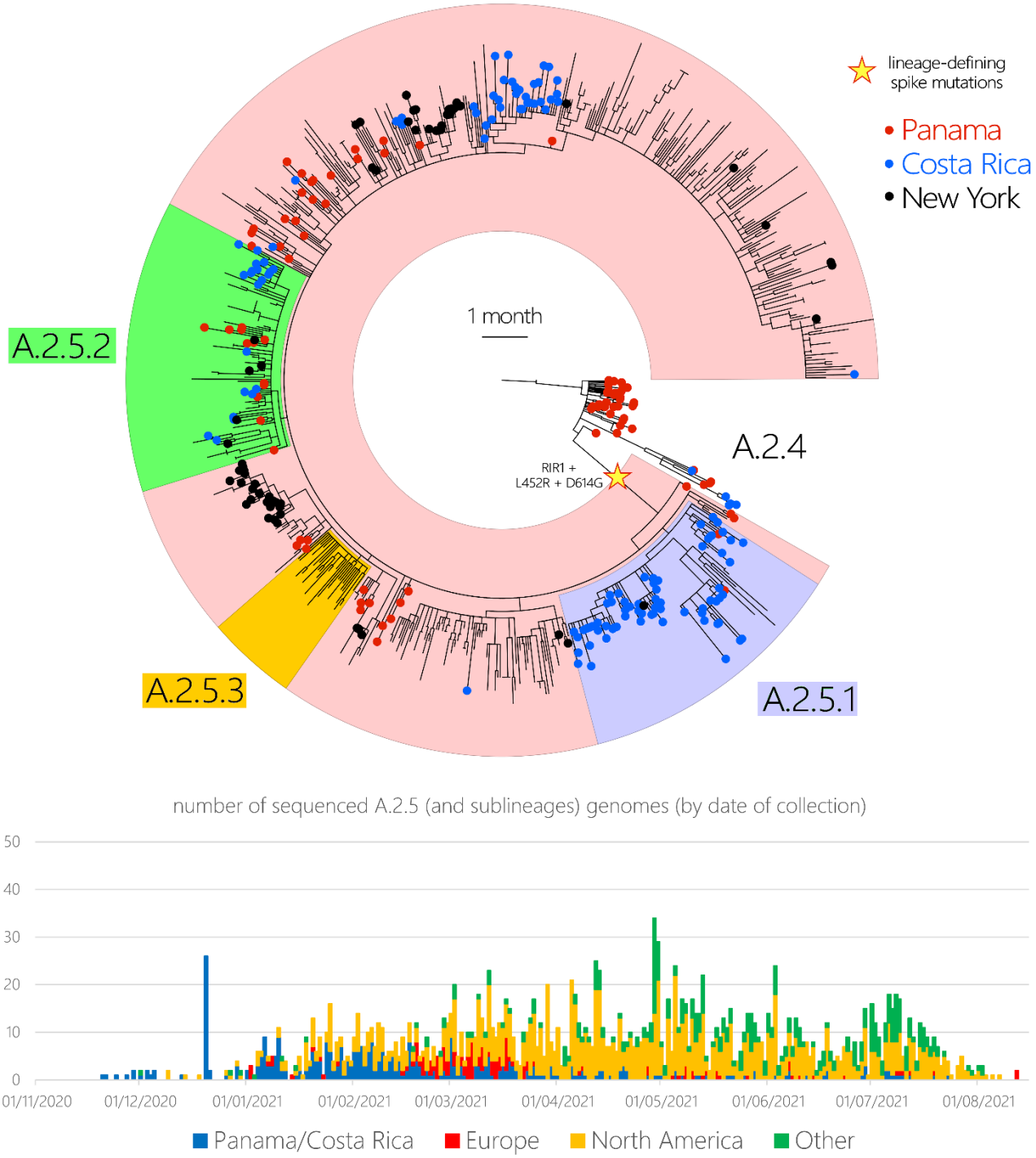


Figure 4. Upper panel: circular time tree exemplifying the phylogeny of the A.2.5 lineage and of its sublineages A.2.5.1 and A.2.5.2. Only high quality, complete genomes have been included. Red, blue and black tip marks indicate samples from Panama, Costa Rica and New York, respectively. The Wuhan-Hu-1 strain was used to root the tree. Lower panel: number of A.2.5 SARS-CoV-2 genomes sequenced in Panama or Costa Rica, Europe, North America and other locations, shown by date of sampling, starting from November 1st 2020. Note that the numbers referred to recent dates are unreliable due to delays in data submission.

Context for the emergence of B.1.214.2

B.1.214.2 is linked with the parent lineage B.1.214, which spread in the early phases of the pandemics in the Democratic Republic of Congo (Ntoumi et al., 2021) and possibly in other neighboring countries where limited genomic surveillance was carried out in 2020. As a result of exportations, four distinct sister clades later spread, with different success, in Europe between December 2020 and the first months of 2021: B.1.214.1, the most basal of these and the only one to lack the T617I spike mutation, B.1.214.2, B.1.214.3 and B.1.214.4. Out of the four sister lineages, B.1.214.2 became, by far, the most widespread in Europe in early 2021 (**Figure 5**). The very first cases linked with this lineage in the continent were detected in late November 2020 in Switzerland, and more specifically in the Basel-Landschaft. This introduction led to continued community transmission, as evidenced by the large monophyletic cluster of genotypes visible in the phylogenetic tree (**Figure 5**). The circulation of B.1.214.2 continued to be significant in the Basel region in early 2021 (e.g. its estimated frequency reached 8.5% in March 2021), but progressively declined along with the spread of alpha, and no B.1.214.2 isolate has been sequenced in Switzerland after May 20th, 2021. In Europe, B.1.214.2 found its most significant spread in Belgium, where it was first detected on January 3rd 2021 in Liège, and underwent a rapid increase in frequency, accounting for 5-6% of all the genomes sequenced between the 7th and 9th epidemiological weeks of surveillance. Phylogeny clearly indicates that this rapid spread was linked with a higher genomic heterogeneity compared with Switzerland, suggesting the possibility of multiple independent introductions from abroad (**Figure 5**). The main candidate as the source for such undocumented introduction is Central Africa, taking into account the geographical origins of the parent lineage B.1.214, the early detection of B.1.214.2 in the Democratic Republic of Congo (i.e. January 2021) and the historical connections between the two countries. As in the case of Switzerland, the frequency of observation of B.1.214.2 underwent a sharp decline, in parallel with the rise of alpha. The subsequent spread of delta led to the apparent disappearance of B.1.614.2 from all Europe, with the most recent detection of genotypes belonging to this lineage dated June 28th, 2021 (in Begium).

The large Belgian cluster of community transmission led to the exportation of cases in a number of other European countries, to the point that B.1.214.2 was identified as a variant under monitoring by both Public Health England (Public Health England, 2021) and ECDC (https://www.ecdc.europa.eu/en/covid-19/variants-concern). Besides Europe, in addition to the previously mentioned cases reported in the Democratic Republic of Congo, some infections have been reported in other African countries (Congo, Angola, Benin, Gabon, Rwanda, Togo and Senegal) and in Asia (India, Indonesia and Malaysia). Moreover, several B.1.215.2 genomes have been sequenced across the United States, mostly in Colorado, Maryland and Virginia, indicating multiple introductions and signs of community transmission. Nevertheless, as previously reported for Europe, the number of sequenced B.1.214.2 genomes started to decrease globally along with the spread of alpha and further declined when delta became prevalent (**Figure 5**). Based on the lack of detection over the past two months in large-scale sequencing-based surveillance efforts (the latest detection dates July 7th, 2021 in Rwanda), B.1.214.2 might be provisionally considered as an extinct SARS-CoV-2 lineage.

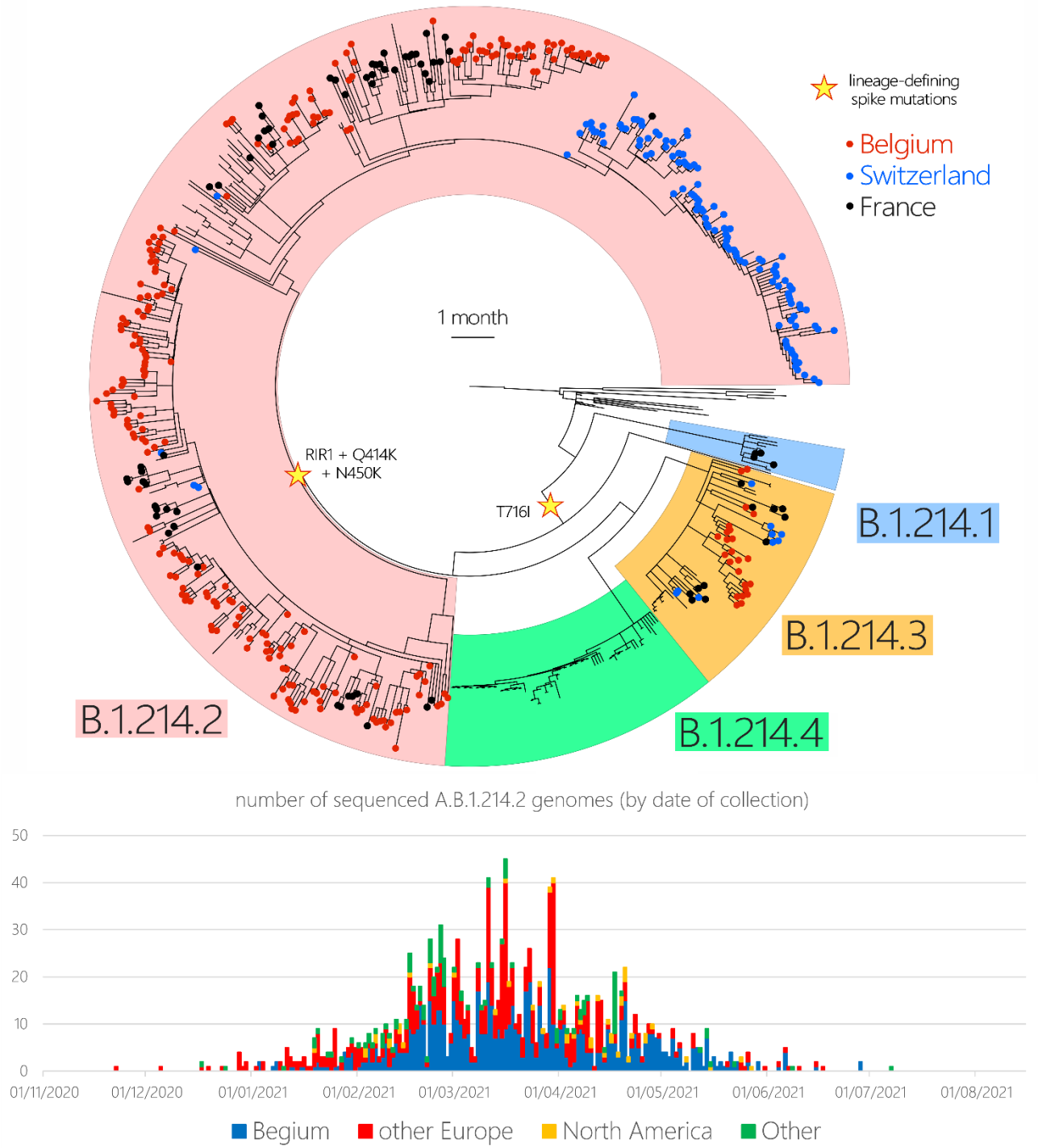


Figure 5. Upper panel: circular time tree exemplifying the phylogeny of the B.1.214.2 lineage, of the parental lineage B.1.214, and of the sister lineages B.1.214.1, B.1.214.3 and B.1.214.4. Only high quality, complete genomes have been included. Red, blue and black tip marks indicate samples from Belgium, Switzerland and France, respectively. The Wuhan-Hu-1 strain was used to root the tree. Lower panel: number of B.1.214.2 SARS-CoV-2 genomes sequenced in Belgium, other European countries, North America and other locations, shown by date of sampling, starting from November 1st 2020. Note that the numbers referred to recent dates are unreliable due to delays in data submission.

Structural considerations

RDR1 is located in a loop which connects the spike NTD β strands 15 and 16, a region which, unlike RDR1, 2, 3 and 4, does not show any overlap with any known major NTD antigenic sites (Cerutti et al., 2021; McCallum

et al., 2021). Hence, the involvement of the insertions reported in this manuscript in antibody escape is unlikely, even though the possibility that this modification may lead to paired structural alterations at distantly related sites, leading to a reduced surface accessibility of canonical antibody epitopes cannot be ruled out. Interestingly, previous comparative genomics investigations carried out in the Sarbecovirus subgenus revealed that the spike protein of both RmYN02 and RacCS203 (Zhou et al., 2021, 2020), which are the closest known relatives to SARS-CoV-2 when genomic recombination is taken into account (Lytras et al., 2021), comprises an insertion of four codons at RIR1 in comparison with other Sarbecoviruses. This observation points out that RIR1 has been previously prone to structural alterations during the radiation of bat coronaviruses (Garry et al., 2021). Although the reasons behind the recurrent occurrence of insertions at RIR1, other authors have previously reported that indels in the SARS-CoV-2 genome are likely to be the product of transcription regulatory site-independent template switching, which is apparently driven by secondary structure rearrangements of the folded viral RNA (Chrisman et al., 2021).

Most certainly, the spread of the A.2.5 and B.1.214.2 lineages in different geographical contexts between late 2020 and early 2021 suggests that RIR1 insertions are unlikely to have a detrimental impact on the three-dimensional structure of the spike protein or to dramatically reduce the infectivity of these variants compared with other lineages widespread at that time. At the same time, the well-defined length of the insertions (in the overwhelming majority of cases 3 or 4 codons) suggests that some critical structural constraints, that may prevent the selection of shorter insertions or limit their associated evolutionary benefits, might exist. Several spike mutations located in the NTD can affect the structural organization of the spike protein, altering the stability of the interaction between the RBD and the ACE2 receptor, or its accessibility to antibody recognition. For instance, the NTD Δ69/Δ70 deletion, which, like RIR1, is found in multiple independent lineages, does not determine a significant antibody escape *in vitro* (P. Wang et al., 2021). However, it is thought to have an important impact on the structure of the spike protein, by conferring increased cleavage at the S1/S2 site, thereby allowing higher incorporation in the virion (Kemp et al., 2021). In light of these observations, some NTD indels apparently not related with immune escape may act as permissive mutations, by compensating small infectivity deficits associated with other RBD mutations (i.e. L452R in A.2.5 and Q414K and N450K in B.1.214.2).

To preliminarily investigate the impact of RIR1 insertions on the structure of the spike protein, we applied molecular dynamics simulations, a well-known technique able to capture and study the dynamical properties of proteins and to assess the effects of mutations, deletions and insertions (Hansson et al., 2002). In this case, we have used a coarse-grained force-field to compare the structural and/or dynamical differences between the spike proteins from the wild-type virus and those from the A.2.5 and B.1.214.2 lineages. After 2µs microseconds of simulations, the RMSD of the backbone atoms relative to the equivalent initial structures (which represents a global measure of protein fluctuations) was calculated as a function of time to evaluate the stability of MD simulations equilibrium in the three systems. No significant global displacement was detected for any of the three protein models compared with the initial structure, as most of the RMSD values only displayed fluctuations in a range between 0.35Å and 0.45Å. Similarly, the presence of a few spike mutations in A.2.5 and B.1.214.2 only led to minor changes in the compactness of the proteins, as suggested by the differences of about 1 Å found in the average RYGR values among the three models (**Supplementary Figure 1**). On the other hand, some fluctuations were visible in the RMSF of the A.2.5 and B.1.214.2 spike protein models, in particular in the regions which harbored non-synonymous mutations compared with the wild-type protein.

To understand changes in the direction of motions of the different systems under analysis, PCA was performed on the last 2µs of the simulations, the time after which the systems reached the equilibration state. The analysis was then restricted to the backbone beads, as they are less perturbed by statistical noise, providing at the same time a significant characterization of the essential space motions (Michaud-Agrawal et

al., 2011). The diagonalization of the covariance matrix of fluctuations of the residues belonging to the backbone resulted in a set of eigenvalues, which were plotted in decreasing order against the corresponding eigenvector indices. The first few eigenvectors corresponded to concerted motions that quickly decreased in amplitude to reach some constrained and more localized fluctuations. Here we present the principal modes along the first eigenvector (**Figure 6 A**), which covers about 25% of the motions of the protein. Consistently with the placement of non-synonymous mutations (**Figure 3**), this analysis revealed that both A.2.5 and B.1.214.2 exhibited some changes in the fluctuations in regions belonging to the NTD and RBD, which are shown in red in **Figure 6 B**. This indicates that the presence of the mutations and insertions may induce local structural and dynamical changes on the spike protein, highlighting the usefulness of performing studies on the dynamical properties of insertions upon their emergence in variants with widespread circulation.

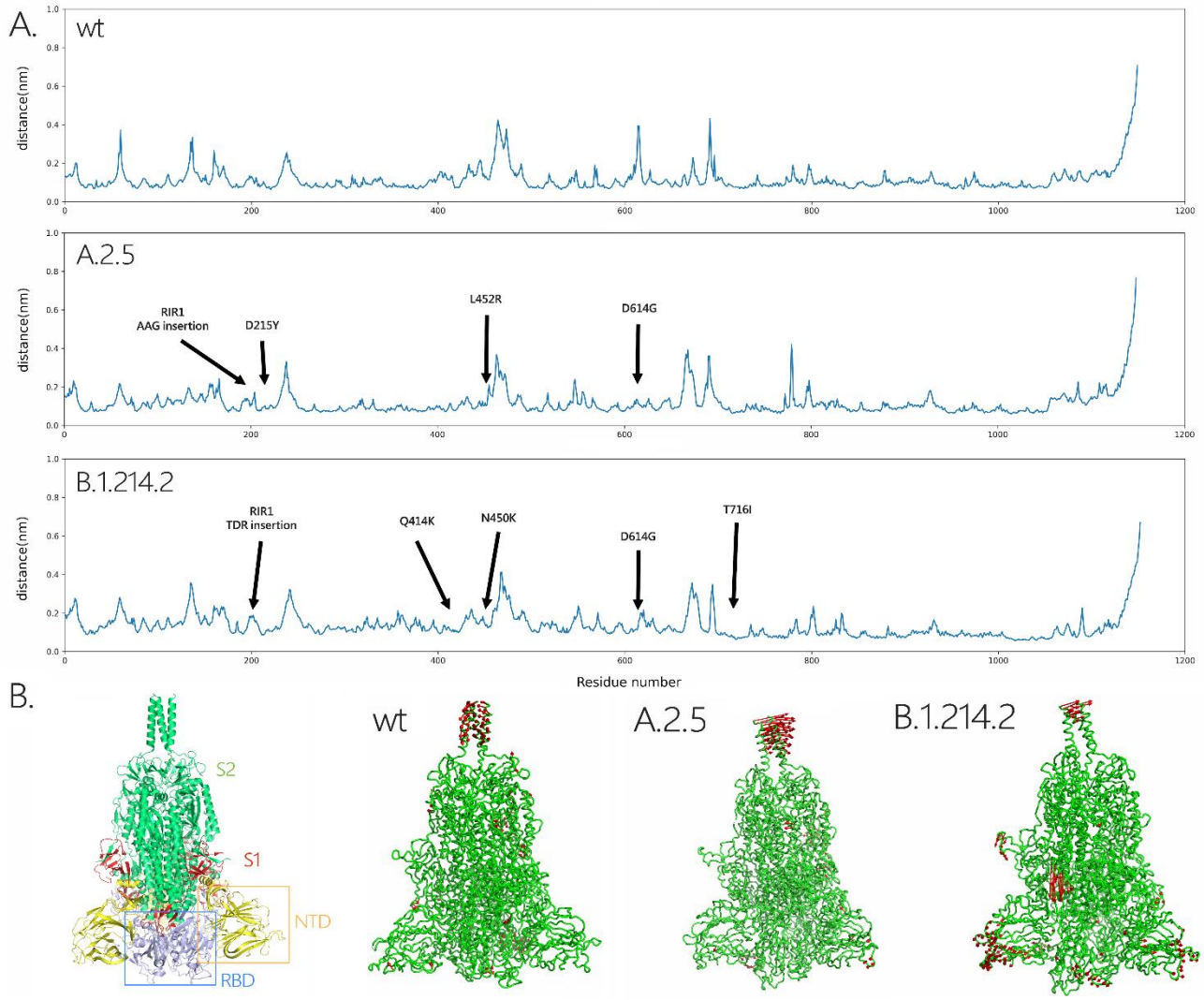


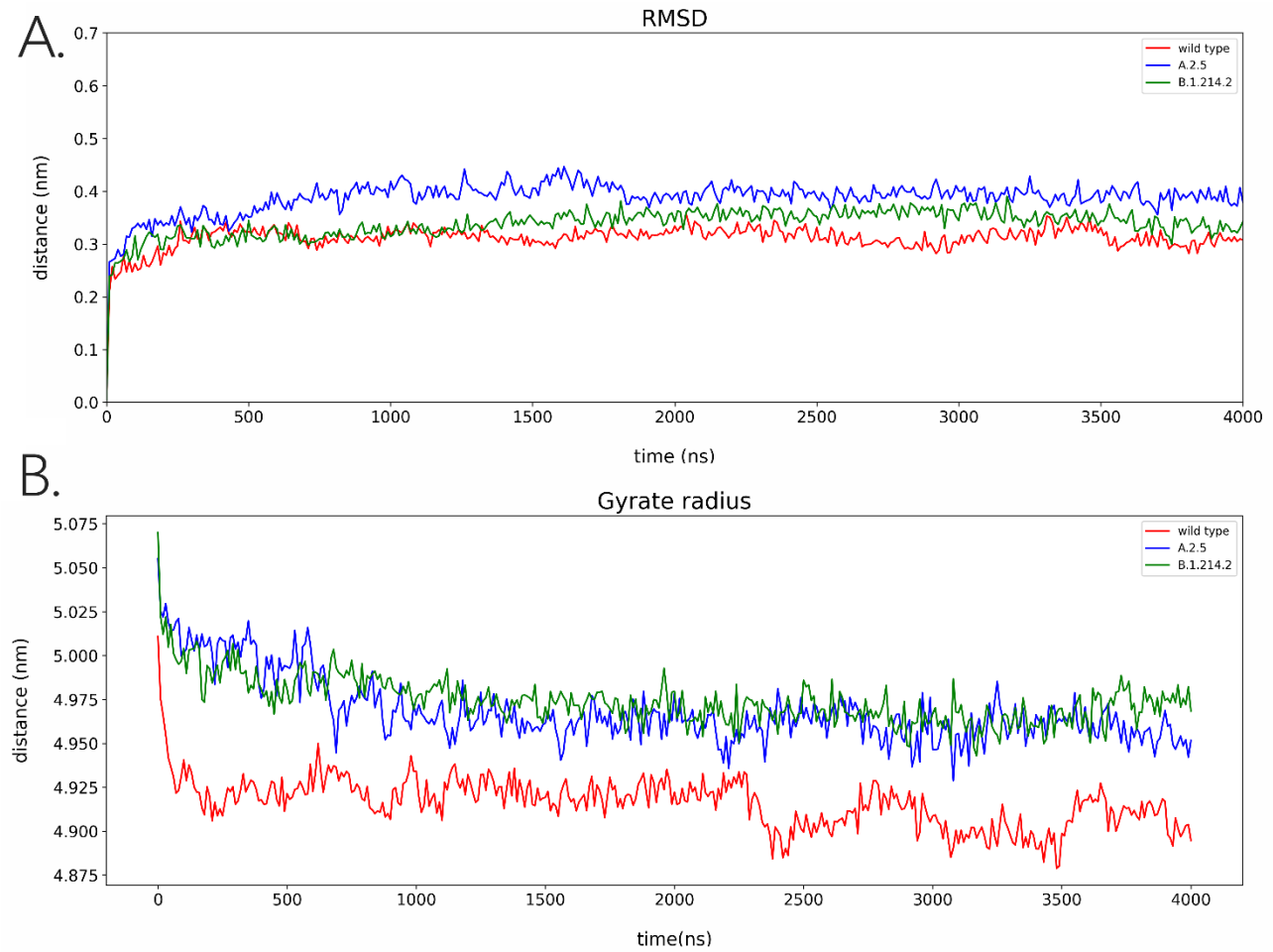
Figure 6: Panel A: RMSF plot for the models of the SARS-CoV-2 spike protein (wild-type, A.2.5 and B.1.214.2), with indication of the point and insertion mutations present in the two viral lineages target of his study, compared with the wild type virus. Panel B: Three-dimensional structural models obtained for the wild type spike protein and the A.2.5 and B.1.214.2 variants. The location of the NTD and RBD (within the S1 subunit) and of the S2 subunit in the spike trimer are shown at the left-hand side. The regions where the most significant fluctuations were evidenced are marked in red.

Conclusions

The SARS-CoV-2 genome continues to accumulate mutations at a relatively constant rate, occasionally originating new VOCs and VOIs as a result of continued high viral circulation and natural selection. The insertions at RIR1 documented in this work have only led, so far, to the emergence of two viral lineages with worldwide distribution, and one of these (B.1.214.2) can be considered as provisionally extinct. However, the recurrent independent occurrence of this phenomenon by convergent evolution in multiple viral lineages suggests that it may be linked with an evolutionary advantage, whose magnitude is presently unclear. Since insertions at RIR have already been documented in some VOCs and VOIs, including delta, our observations suggest that the possible emergence of similar mutations in quickly spreading variants should be closely monitored in the future.

In absence of functional data, the role of RIR1 insertions can be only speculated. However, based on the lack of overlap with known immune epitopes their involvement in immune escape phenomena appears unlikely. On the other hand, we observe a correlation between the presence of RIR1 insertions and non-synonymous mutations found in the RBD. This, together with the predicted impact on the structure of the spike protein, may suggest a possible role as a permissive mutation in compensating otherwise disadvantageous non-synonymous spike mutations.

Supplementary Figures



Supplementary Figure 1: Panel A: RMSD plot of the three models of the SARS-CoV-2 spike protein (wild-type, A.2.5 and B.1.214.2), as a function of simulated time. The systems reach equilibrium after 1.5 µs. Panel B: variation of RGYR observed over time for the three models of the SARS-CoV-2 spike protein (wild-type, A.2.5 and B.1.214.2) during the MD simulation.

Acknowledgements

The author is grateful to Dr. Alberto Beretta, Dr. Sarah Ann Nadeau and Dr. Alexander Martinez for their communication and useful suggestions, and the staff of Pop Medicine (https://www.facebook.com/medicipop/) for the support. The author also acknowledges all the fundamental work carried out by the clinicians, researchers and public health authorities that allowed the collection of SARS-CoV-2 genome data and made sequence data available in a timely manner though GISAID (Shu and McCauley, 2017), as well as the great efforts made by the developers of nextstrain to assist researchers in SARS-CoV-2 evolution studies.

Competing Interest Statement

The author declares he has no competing interests.

Funding Statement

The authors received no funding for this work.

References

Aguayo-Ortiz, R., Chávez-García, C., Straub, J.E., Dominguez, L., 2017. Characterizing the structural ensemble of γ-secretase using a multiscale molecular dynamics approach. Chem. Sci. 8, 5576–5584. https://doi.org/10.1039/C7SC00980A

Alouane, T., Laamarti, M., Essabbar, A., Hakmi, M., Bouricha, E.M., Chemao-Elfihri, M.W., Kartti, S., Boumajdi, N., Bendani, H., Laamarti, R., Ghrifi, F., Allam, L., Aanniz, T., Ouadghiri, M., El Hafidi, N., El Jaoudi, R., Benrahma, H., Attar, J.E., Mentag, R., Sbabou, L., Nejjari, C., Amzazi, S., Belyamani, L., Ibrahimi, A., 2020. Genomic Diversity and Hotspot Mutations in 30,983 SARS-CoV-2 Genomes: Moving Toward a Universal Vaccine for the "Confined Virus"? Pathog. Basel Switz. 9. https://doi.org/10.3390/pathogens9100829

Bernal, J.L., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., Stowe, J., Tessier, E., Groves, N., Dabrera, G., Myers, R., Campbell, C., Amirthalingam, G., Edmunds, M., Zambon, M., Brown, K., Hopkins, S., Chand, M., Ramsay, M., 2021. Effectiveness of COVID-19 vaccines against the B.1.617.2 variant. medRxiv 2021.05.22.21257658. https://doi.org/10.1101/2021.05.22.21257658

Boni, M.F., Lemey, P., Jiang, X., Lam, T.T.-Y., Perry, B.W., Castoe, T.A., Rambaut, A., Robertson, D.L., 2020. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. 5, 1408–1417. https://doi.org/10.1038/s41564-020-0771-4

Bugembe, D.L., Phan, M.V.T., Ssewanyana, I., Semanda, P., Nansumba, H., Dhaala, B., Nabadda, S., O'Toole, Á.N., Rambaut, A., Kaleebu, P., Cotten, M., 2021. A SARS-CoV-2 lineage A variant (A.23.1) with altered spike has emerged and is dominating the current Uganda epidemic. medRxiv 2021.02.08.21251393. https://doi.org/10.1101/2021.02.08.21251393

Bussi, G., Donadio, D., Parrinello, M., 2007. Canonical sampling through velocity rescaling. J. Chem. Phys. 126, 014101. https://doi.org/10.1063/1.2408420

Cai, Y., Zhang, J., Xiao, T., Peng, H., Sterling, S.M., Walsh, R.M., Rawson, S., Rits-Volloch, S., Chen, B., 2020. Distinct conformational states of SARS-CoV-2 spike protein. Science 369, 1586–1592. https://doi.org/10.1126/science.abd4251

Cerutti, G., Guo, Y., Zhou, T., Gorman, J., Lee, M., Rapp, M., Reddem, E.R., Yu, J., Bahna, F., Bimela, J., Huang, Y., Katsamba, P.S., Liu, L., Nair, M.S., Rawi, R., Olia, A.S., Wang, P., Zhang, B., Chuang, G.-Y., Ho, D.D., Sheng, Z., Kwong, P.D., Shapiro, L., 2021. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe. https://doi.org/10.1016/j.chom.2021.03.005

Choi, B., Choudhary, M.C., Regan, J., Sparks, J.A., Padera, R.F., Qiu, X., Solomon, I.H., Kuo, H.-H., Boucau, J., Bowman, K., Adhikari, U.D., Winkler, M.L., Mueller, A.A., Hsu, T.Y.-T., Desjardins, M., Baden, L.R., Chan, B.T., Walker, B.D., Lichterfeld, M., Brigl, M., Kwon, D.S., Kanjilal, S., Richardson, E.T., Jonsson, A.H., Alter, G., Barczak, A.K., Hanage, W.P., Yu, X.G., Gaiha, G.D., Seaman, M.S., Cernadas, M., Li, J.Z., 2020. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N. Engl. J. Med. 383, 2291–2293. https://doi.org/10.1056/NEJMc2031364

Chrisman, B.S., Paskov, K., Stockham, Nate., Tabatabaei, K., Jung, J.-Y., Washington, P., Varma, M., Sun, M.W., Maleki, S., Wall, D.P., 2021. Indels in SARS-CoV-2 occur at template-switching hotspots. BioData Min. 14, 20. https://doi.org/10.1186/s13040-021-00251-0

Dagpunar, J., 2021. Interim estimates of increased transmissibility, growth rate, and reproduction number of the Covid-19 B.1.617.2 variant of concern in the United Kingdom. medRxiv 2021.06.03.21258293. https://doi.org/10.1101/2021.06.03.21258293

Davies, N.G., Abbott, S., Barnard, R.C., Jarvis, C.I., Kucharski, A.J., Munday, J.D., Pearson, C.A.B., Russell, T.W., Tully, D.C., Washburne, A.D., Wenseleers, T., Gimma, A., Waites, W., Wong, K.L.M., Zandvoort, K. van, Silverman, J.D., Group1‡, C.C.-19 W., Consortium‡, C.-19 G.U. (COG-U., Diaz-Ordaz, K., Keogh,

R., Eggo, R.M., Funk, S., Jit, M., Atkins, K.E., Edmunds, W.J., 2021. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. Science. https://doi.org/10.1126/science.abg3055

DeLano, W.L., 2002. Pymol: An open-source molecular graphics tool. CCP4 Newsl. Protein Crystallogr. 40, 82–92.

Denison, M.R., Graham, R.L., Donaldson, E.F., Eckerle, L.D., Baric, R.S., 2011. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. 8, 270–279. https://doi.org/10.4161/rna.8.2.15013

Dhar, M.S., Marwal, R., Radhakrishnan, V.S., Ponnusamy, K., Jolly, B., Bhoyar, R.C., Fatihi, S., Datta, M., Singh, P., Sharma, U., Ujjainia, R., Naushin, S., Bhateja, N., Divakar, M.K., Sardana, V., Singh, M.K., Imran, M., Senthivel, V., Maurya, R., Jha, N., Mehta, P., Rophina, M., Arvinden, V.R., Chaudhary, U., Thukral, L., Pandey, R., Dash, D., Faruq, M., Lall, H., Gogia, H., Madan, P., Kulkarni, S., Chauhan, H., Sengupta, S., Kabra, S., Consortium (INSACOG), T.I.S.-C.-2 G., Singh, S.K., Agrawal, A., Rakshit, P., 2021. Genomic characterization and Epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. medRxiv 2021.06.02.21258076. https://doi.org/10.1101/2021.06.02.21258076

Franco, D., Gonzalez, C., Abrego, L.E., Carrera, J.-P., Diaz, Y., Caicedo, Y., Moreno, A., Chavarria, O., Gondola, J., Castillo, M., Valdespino, E., Gaitán, M., Martínez-Mandiche, J., Hayer, L., Gonzalez, P., Lange, C., Molto, Y., Mojica, D., Ramos, R., Mastelari, M., Cerezo, L., Moreno, L., Donnelly, C.A., Pascale, J.M., Faria, N.R., Lopez-Verges, S., Martinez, A.A., 2020. Early Transmission Dynamics, Spread, and Genomic Characterization of SARS-CoV-2 in Panama. Emerg. Infect. Dis. 27, 612–615. https://doi.org/10.3201/eid2702.203767

Garry, R.F., Andersen, K.G., Gallaher, W.R., Tsan-Yuk Lam, T., Gangaparapu, K., Latif, A.A., Beddingfield, B.J., Rambaut, A., Holmes, E.C., 2021. Spike protein mutations in novel SARS-CoV-2 'variants of concern' commonly occur in or near indels [WWW Document]. Virological. URL https://virological.org/t/spike-protein-mutations-in-novel-sars-cov-2-variants-of-concern-commonly-occur-in-or-near-indels/605 (accessed 5.5.21).

Ge, X.-Y., Wang, N., Zhang, W., Hu, B., Li, B., Zhang, Y.-Z., Zhou, J.-H., Luo, C.-M., Yang, X.-L., Wu, L.-J., Wang, B., Zhang, Y., Li, Z.-X., Shi, Z.-L., 2016. Coexistence of multiple coronaviruses in several bat colonies in an abandoned mineshaft. Virol. Sin. 31, 31–40. https://doi.org/10.1007/s12250-016-3713-9

Greaney, A.J., Starr, T.N., Barnes, C.O., Weisblum, Y., Schmidt, F., Caskey, M., Gaebler, C., Cho, A., Agudelo, M., Finkin, S., Wang, Z., Poston, D., Muecksch, F., Hatziioannou, T., Bieniasz, P.D., Robbiani, D.F., Nussenzweig, M.C., Bjorkman, P.J., Bloom, J.D., 2021. Mutational escape from the polyclonal antibody response to SARS-CoV-2 infection is largely shaped by a single class of antibodies. bioRxiv 2021.03.17.435863. https://doi.org/10.1101/2021.03.17.435863

Hansson, T., Oostenbrink, C., van Gunsteren, W., 2002. Molecular dynamics simulations. Curr. Opin. Struct. Biol. 12, 190–196. https://doi.org/10.1016/s0959-440x(02)00308-1

Ignatieva, A., Hein, J., Jenkins, P.A., 2021. Evidence of ongoing recombination in SARS-CoV-2 through genealogical reconstruction. bioRxiv 2021.01.21.427579. https://doi.org/10.1101/2021.01.21.427579

Johnson, B.A., Xie, X., Bailey, A.L., Kalveram, B., Lokugamage, K.G., Muruato, A., Zou, J., Zhang, X., Juelich, T., Smith, J.K., Zhang, L., Bopp, N., Schindewolf, C., Vu, M., Vanderheiden, A., Winkler, E.S., Swetnam, D., Plante, J.A., Aguilar, P., Plante, K.S., Popov, V., Lee, B., Weaver, S.C., Suthar, M.S., Routh, A.L., Ren, P., Ku, Z., An, Z., Debbink, K., Diamond, M.S., Shi, P.-Y., Freiberg, A.N., Menachery, V.D., 2021. Loss of furin cleavage site attenuates SARS-CoV-2 pathogenesis. Nature 591, 293–299. https://doi.org/10.1038/s41586-021-03237-4

Katoh, K., Misawa, K., Kuma, K., Miyata, T., 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30, 3059–3066. https://doi.org/10.1093/nar/gkf436

Kemp, S.A., Meng, B., Ferriera, I.A., Datir, R., Harvey, W.T., Papa, G., Lytras, S., Collier, D.A., Mohamed, A., Gallo, G., Thakur, N., Consortium, T.C.-19 G.U. (COG-U., Carabelli, A.M., Kenyon, J.C., Lever, A.M., Marco, A.D., Saliba, C., Culap, K., Cameroni, E., Piccoli, L., Corti, D., James, L.C., Bailey, D., Robertson, D.L., Gupta, R.K., 2021. Recurrent emergence and transmission of a SARS-CoV-2 spike deletion H69/V70. bioRxiv 2020.12.14.422555. https://doi.org/10.1101/2020.12.14.422555

Kistler, K.E., Bedford, T., 2021. Evidence for adaptive evolution in the receptor-binding domain of seasonal coronaviruses OC43 and 229e. eLife 10, e64509. https://doi.org/10.7554/eLife.64509

Klink, G.V., Safina, K.R., Garushyants, S.K., Moldovan, M., Nabieva, E., Consortium, T.C. (Coronavirus R.G.I., Komissarov, A.B., Lioznov, D., Bazykin, G.A., 2021. Spread of endemic SARS-CoV-2 lineages in Russia. medRxiv 2021.05.25.21257695. https://doi.org/10.1101/2021.05.25.21257695

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., de Silva, T.I., Angyal, A., Brown, R.L., Carrilero, L., Green, L.R., Groves, D.C., Johnson, K.J., Keeley, A.J., Lindsey, B.B., Parsons, P.J., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R.M., Wang, D., Wyles, M.D., McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182, 812-827.e19. https://doi.org/10.1016/j.cell.2020.06.043

Laiton-Donato, K., Franco-Muñoz, C., Álvarez-Díaz, D.A., Ruiz-Moreno, H.A., Usme-Ciro, J.A., Prada, D.A., Reales-González, J., Corchuelo, S., Herrera-Sepúlveda, M.T., Naizaque, J., Santamaría, G., Rivera, J., Rojas, P., Ortiz, J.H., Cardona, A., Malo, D., Prieto-Alvarado, F., Gómez, F.R., Wiesner, M., Martínez, M.L.O., Mercado-Reyes, M., 2021. Characterization of the emerging B.1.621 variant of interest of SARS-CoV-2. medRxiv 2021.05.08.21256619. https://doi.org/10.1101/2021.05.08.21256619

Liu, Z., VanBlargan, L.A., Bloyet, L.-M., Rothlauf, P.W., Chen, R.E., Stumpf, S., Zhao, H., Errico, J.M., Theel, E.S., Liebeskind, M.J., Alford, B., Buchser, W.J., Ellebedy, A.H., Fremont, D.H., Diamond, M.S., Whelan, S.P.J., 2021. Landscape analysis of escape variants identifies SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. bioRxiv 2020.11.06.372037. https://doi.org/10.1101/2020.11.06.372037

Lustig, Y., Nemet, I., Kliker, L., Zuckerman, N., Yishai, R., Alroy-Preis, S., Mendelson, E., Mandelboim, M., 2021. Neutralizing Response against Variants after SARS-CoV-2 Infection and One Dose of BNT162b2. N. Engl. J. Med. 0, null. https://doi.org/10.1056/NEJMc2104036

Lytras, S., Hughes, J., Martin, D., Klerk, A. de, Lourens, R., Pond, S.L.K., Xia, W., Jiang, X., Robertson, D.L., 2021. Exploring the natural origins of SARS-CoV-2 in the light of recombination. bioRxiv 2021.01.22.427830. https://doi.org/10.1101/2021.01.22.427830

Ma, Y., Wu, L., Shaw, N., Gao, Y., Wang, J., Sun, Y., Lou, Z., Yan, L., Zhang, R., Rao, Z., 2015. Structural basis and functional analysis of the SARS coronavirus nsp14–nsp10 complex. Proc. Natl. Acad. Sci. 112, 9436–9441. https://doi.org/10.1073/pnas.1508686112

Madhi, S.A., Baillie, V., Cutland, C.L., Voysey, M., Koen, A.L., Fairlie, L., Padayachee, S.D., Dheda, K., Barnabas, S.L., Bhorat, Q.E., Briner, C., Kwatra, G., Ahmed, K., Aley, P., Bhikha, S., Bhiman, J.N., Bhorat, A.E., du Plessis, J., Esmail, A., Groenewald, M., Horne, E., Hwa, S.-H., Jose, A., Lambe, T., Laubscher, M., Malahleha, M., Masenya, M., Masilela, M., McKenzie, S., Molapo, K., Moultrie, A., Oelofse, S., Patel, F., Pillay, S., Rhead, S., Rodel, H., Rossouw, L., Taoushanis, C., Tegally, H., Thombrayil, A., van Eck, S., Wibmer, C.K., Durham, N.M., Kelly, E.J., Villafana, T.L., Gilbert, S., Pollard, A.J., de Oliveira, T., Moore, P.L., Sigal, A., Izu, A., 2021. Efficacy of the ChAdOx1 nCoV-19 Covid-19 Vaccine against the B.1.351 Variant. N. Engl. J. Med. 0, null. https://doi.org/10.1056/NEJMoa2102214

Martonák, R., Laio, A., Parrinello, M., 2003. Predicting crystal structures: the Parrinello-Rahman method revisited. Phys. Rev. Lett. 90, 075503. https://doi.org/10.1103/PhysRevLett.90.075503

McCallum, M., Marco, A.D., Lempp, F., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., Zepeda, S., Iulio, J. di, Bowen, J.E., Montiel-Ruiz, M., Zhou, J., Rosen, L.E., Bianchi, S., Guarino, B., Fregni, C.S., Abdelnabi, R., Foo, S.-Y.C., Rothlauf, P.W., Bloyet, L.-M., Benigni, F., Cameroni, E., Neyts, J., Riva, A., Snell, G., Telenti, A., Whelan, S.P.J., Virgin, H.W., Corti, D., Pizzuto, M.S., Veesler, D., 2021. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. bioRxiv 2021.01.14.426475. https://doi.org/10.1101/2021.01.14.426475

McCarthy, K.R., Rennick, L.J., Nambulli, S., Robinson-McCarthy, L.R., Bain, W.G., Haidar, G., Duprex, W.P., 2021. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science 371, 1139–1142. https://doi.org/10.1126/science.abf6950

Michaud-Agrawal, N., Denning, E.J., Woolf, T.B., Beckstein, O., 2011. MDAnalysis: a toolkit for the analysis of molecular dynamics simulations. J. Comput. Chem. 32, 2319–2327. https://doi.org/10.1002/jcc.21787

Millet, J.K., Jaimes, J.A., Whittaker, G.R., 2020. Molecular diversity of coronavirus host cell entry receptors. FEMS Microbiol. Rev. https://doi.org/10.1093/femsre/fuaa057

Molina-Mora, J.A., Cordero-Laurent, E., Godínez, A., Calderón, M., Brenes, H., Soto-Garita, C., Pérez-Corrales, C., Rica, C.-C.C.I. de E.G. del S.-C.-2 C., Drexler, J.F., Moreira-Soto, A., Corrales-Aguilar, E., Duarte-Martínez, F., 2020. SARS-CoV-2 Genomic Surveillance in Costa Rica: Evidence of a Divergent Population and an Increased Detection of a Spike T1117I Mutation. bioRxiv 2020.12.21.423850. https://doi.org/10.1101/2020.12.21.423850

Nelson, G., Buzko, O., Spilman, P., Niazi, K., Rabizadeh, S., Soon-Shiong, P., 2021. Molecular dynamic simulation reveals E484K mutation enhances spike RBD-ACE2 affinity and the combination of E484K, K417N and N501Y mutations (501Y.V2 variant) induces conformational change greater than N501Y mutant alone, potentially resulting in an escape mutant. bioRxiv 2021.01.13.426558. https://doi.org/10.1101/2021.01.13.426558

Ntoumi, F., Mfoutou Mapanguy, C.C., Tomazatos, A., Pallerla, S.R., Linh, L.T.K., Casadei, N., Angelov, A., Sonnabend, M., Peter, S., Kremsner, P.G., Velavan, T.P., 2021. Genomic surveillance of SARS-CoV-2 in the Republic of Congo. Int. J. Infect. Dis. 105, 735–738. https://doi.org/10.1016/j.ijid.2021.03.036

Ozer, E.A., Simons, L.M., Adewumi, O.M., Fowotade, A.A., Omoruyi, E.C., Adeniji, J.A., Dean, T.J., Taiwo, B.O., Hultquist, J.F., Lorenzo-Redondo, R., 2021. High prevalence of SARS-CoV-2 B.1.1.7 (UK variant) and the novel B.1.525 lineage in Oyo State, Nigeria. medRxiv 2021.04.09.21255206. https://doi.org/10.1101/2021.04.09.21255206

Peacock, T.P., Sheppard, C.M., Brown, J.C., Goonawardane, N., Zhou, J., Whiteley, M., Consortium, P.V., Silva, T.I. de, Barclay, W.S., 2021. The SARS-CoV-2 variants associated with infections in India, B.1.617, show enhanced spike cleavage by furin. bioRxiv 2021.05.28.446163. https://doi.org/10.1101/2021.05.28.446163

Planas, D., Veyer, D., Baidaliuk, A., Staropoli, I., Guivel-Benhassine, F., Rajah, M.M., Planchais, C., Porrot, F., Robillard, N., Puech, J., Prot, M., Gallais, F., Gantner, P., Velay, A., Guen, J.L., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., Péré, H., Courtellemont, L., Hocqueloux, L., Fafi-Kremer, S., Prazuck, T., Mouquet, H., Bruel, T., Simon-Lorière, E., Rey, F.A., Schwartz, O., 2021. Reduced sensitivity of infectious SARS-CoV-2 variant B.1.617.2 to monoclonal antibodies and sera from convalescent and vaccinated individuals. bioRxiv 2021.05.26.445838. https://doi.org/10.1101/2021.05.26.445838

Price, M.N., Dehal, P.S., Arkin, A.P., 2010. FastTree 2 – Approximately Maximum-Likelihood Trees for Large Alignments. PLOS ONE 5, e9490. https://doi.org/10.1371/journal.pone.0009490

Public Health England, 2021. SARS-CoV-2 variants of concern and variants under investigation in England. Technical briefing 18.

Ren, L., Zhang, Y., Li, J., Xiao, Y., Zhang, J., Wang, Y., Chen, L., Paranhos-Baccalà, G., Wang, J., 2015. Genetic drift of human coronavirus OC43 spike gene during adaptive evolution. Sci. Rep. 5. https://doi.org/10.1038/srep11451

Ren, W., Qu, X., Li, W., Han, Z., Yu, M., Zhou, P., Zhang, S.-Y., Wang, L.-F., Deng, H., Shi, Z., 2008. Difference in Receptor Usage between Severe Acute Respiratory Syndrome (SARS) Coronavirus and SARS-Like Coronavirus of Bat Origin. J. Virol. 82, 1899–1907. https://doi.org/10.1128/JVI.01085-07

Resende, P.C., Naveca, F.G., Lins, R.D., Dezordi, F.Z., Ferraz, M.V.F., Moreira, E.G., Coêlho, D.F., Motta, F.C., Paixão, A.C.D., Appolinario, L., Lopes, R.S., Mendonça, A.C. da F., Rocha, A.S.B. da, Nascimento, V., Souza, V., Silva, G., Nascimento, F., Neto, L.G.L., Riediger, I., Debur, M. do C., Leite, A.B., Mattos, T., Costa, C.F. da, Pereira, F.M., Khouri, R., Bernardes, A.F.L., Delatorre, E., Gräf, T., Siqueira, M.M., Bello, G., Wallau, G.L., 2021. The ongoing evolution of variants of concern and interest of SARS-CoV-2 in Brazil revealed by convergent indels in the amino (N)-terminal domain of the Spike protein. medRxiv 2021.03.19.21253946. https://doi.org/10.1101/2021.03.19.21253946

Sabino, E.C., Buss, L.F., Carvalho, M.P.S., Prete, C.A., Crispim, M.A.E., Fraiji, N.A., Pereira, R.H.M., Parag, K.V., Peixoto, P. da S., Kraemer, M.U.G., Oikawa, M.K., Salomon, T., Cucunuba, Z.M., Castro, M.C., Santos, A.A. de S., Nascimento, V.H., Pereira, H.S., Ferguson, N.M., Pybus, O.G., Kucharski, A., Busch, M.P., Dye, C., Faria, N.R., 2021. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. The Lancet 397, 452–455. https://doi.org/10.1016/S0140-6736(21)00183-5

Sagulenko, P., Puller, V., Neher, R.A., 2018. TreeTime: Maximum-likelihood phylodynamic analysis. Virus Evol. 4. https://doi.org/10.1093/ve/vex042

Shen, X., Tang, H., Pajon, R., Smith, G., Glenn, G.M., Shi, W., Korber, B., Montefiori, D.C., 2021. Neutralization of SARS-CoV-2 Variants B.1.429 and B.1.351. N. Engl. J. Med. 0, null. https://doi.org/10.1056/NEJMc2103740

Shu, Y., McCauley, J., 2017. GISAID: Global initiative on sharing all influenza data – from vision to reality. Eurosurveillance 22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494

Singh, A., Steinkellner, G., Köchl, K., Gruber, K., Gruber, C.C., 2021. Serine 477 plays a crucial role in the interaction of the SARS-CoV-2 spike protein with the human receptor ACE2. Sci. Rep. 11, 4320. https://doi.org/10.1038/s41598-021-83761-5

Starr, T.N., Greaney, A.J., Dingens, A.S., Bloom, J.D., 2021a. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. bioRxiv 2021.02.17.431683. https://doi.org/10.1101/2021.02.17.431683

Starr, T.N., Greaney, A.J., Dingens, A.S., Bloom, J.D., 2021b. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. Cell Rep. Med. 2. https://doi.org/10.1016/j.xcrm.2021.100255

Starr, T.N., Greaney, A.J., Hilton, S.K., Ellis, D., Crawford, K.H.D., Dingens, A.S., Navarro, M.J., Bowen, J.E., Tortorici, M.A., Walls, A.C., King, N.P., Veesler, D., Bloom, J.D., 2020. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. Cell 182, 1295-1310.e20. https://doi.org/10.1016/j.cell.2020.08.012

Sykes, W., Mhlanga, L., Swanevelder, R., Glatt, T.N., Grebe, E., Coleman, C., Pieterson, N., Cable, R., Welte, A., Berg, K. van den, Vermeulen, M., 2021. Prevalence of anti-SARS-CoV-2 antibodies among blood donors in Northern Cape, KwaZulu-Natal, Eastern Cape, and Free State provinces of South Africa in January 2021. (preprint). In Review. https://doi.org/10.21203/rs.3.rs-233375/v1

Teng, S., Sobitan, A., Rhoades, R., Liu, D., Tang, Q., 2021. Systemic effects of missense mutations on SARS-CoV-2 spike glycoprotein stability and receptor-binding affinity. Brief. Bioinform. 22, 1239–1253. https://doi.org/10.1093/bib/bbaa233

Teruel, N., Mailhot, O., Najmanovich, R.J., 2020. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. bioRxiv 2020.12.16.423118. https://doi.org/10.1101/2020.12.16.423118

Van Der Spoel, D., Lindahl, E., Hess, B., Groenhof, G., Mark, A.E., Berendsen, H.J.C., 2005. GROMACS: fast, flexible, and free. J. Comput. Chem. 26, 1701–1718. https://doi.org/10.1002/jcc.20291

van Dorp, L., Acman, M., Richard, D., Shaw, L.P., Ford, C.E., Ormond, L., Owen, C.J., Pang, J., Tan, C.C.S., Boshier, F.A.T., Ortiz, A.T., Balloux, F., 2020a. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect. Genet. Evol. 83, 104351. https://doi.org/10.1016/j.meegid.2020.104351

van Dorp, L., Richard, D., Tan, C.C.S., Shaw, L.P., Acman, M., Balloux, F., 2020b. No evidence for increased transmissibility from recurrent mutations in SARS-CoV-2. Nat. Commun. 11, 1–8. https://doi.org/10.1038/s41467-020-19818-2

Vavrek, D., Speroni, L., Curnow, K.J., Oberholzer, M., Moeder, V., Febbo, P.G., 2021. Genomic surveillance at scale is required to detect newly emerging strains at an early timepoint. medRxiv 2021.01.12.21249613. https://doi.org/10.1101/2021.01.12.21249613

Volz, E., Mishra, S., Chand, M., Barrett, J.C., Johnson, R., Geidelberg, L., Hinsley, W.R., Laydon, D.J., Dabrera, G., O'Toole, Á., Amato, R., Ragonnet-Cronin, M., Harrison, I., Jackson, B., Ariani, C.V., Boyd, O., Loman, N.J., McCrone, J.T., Gonçalves, S., Jorgensen, D., Myers, R., Hill, V., Jackson, D.K., Gaythorpe, K., Groves, N., Sillitoe, J., Kwiatkowski, D.P., Flaxman, S., Ratmann, O., Bhatt, S., Hopkins, S., Gandy, A., Rambaut, A., Ferguson, N.M., 2021. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. Nature 1–17. https://doi.org/10.1038/s41586-021-03470-x

Wacharapluesadee, S., Tan, C.W., Maneeorn, P., Duengkae, P., Zhu, F., Joyjinda, Y., Kaewpom, T., Chia, W.N., Ampoot, W., Lim, B.L., Worachotsueptrakun, K., Chen, V.C.-W., Sirichan, N., Ruchisrisarod, C., Rodpan, A., Noradechanon, K., Phaichana, T., Jantarat, N., Thongnumchaima, B., Tu, C., Crameri, G., Stokes, M.M., Hemachudha, T., Wang, L.-F., 2021. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. Nat. Commun. 12, 972. https://doi.org/10.1038/s41467-021-21240-1

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281-292.e6. https://doi.org/10.1016/j.cell.2020.02.058

Wang, G.-L., Wang, Z.-Y., Duan, L.-J., Meng, Q.-C., Jiang, M.-D., Cao, J., Yao, L., Zhu, K.-L., Cao, W.-C., Ma, M.-J., 2021. Susceptibility of Circulating SARS-CoV-2 Variants to Neutralization. N. Engl. J. Med. 0, null. https://doi.org/10.1056/NEJMc2103022

Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., Graham, B.S., Mascola, J.R., Chang, J.Y., Yin, M.T., Sobieszczyk, M., Kyratsous, C.A., Shapiro, L., Sheng, Z., Huang, Y., Ho, D.D., 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature 1–6. https://doi.org/10.1038/s41586-021-03398-2

Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finkin, S., Schaefer-Babajew, D., Cipolla, M., Gaebler, C., Lieberman, J.A., Oliveira, T.Y., Yang, Z., Abernathy, M.E., Huey-Tubman, K.E., Hurley, A., Turroja, M., West, K.A., Gordon, K., Millard, K.G., Ramos, V., Silva, J.D., Xu, J., Colbert, R.A., Patel, R., Dizon, J., Unson-O'Brien, C., Shimeliovich, I., Gazumyan, A., Caskey, M., Bjorkman, P.J., Casellas, R., Hatziioannou, T., Bieniasz, P.D., Nussenzweig, M.C., 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. bioRxiv 2021.01.15.426911. https://doi.org/10.1101/2021.01.15.426911

Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P., Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of protein structures and complexes. Nucleic Acids Res. 46, W296–W303. https://doi.org/10.1093/nar/gky427

Xie, X., Zou, J., Fontes-Garfias, C.R., Xia, H., Swanson, K.A., Cutler, M., Cooper, D., Menachery, V.D., Weaver, S., Dormitzer, P.R., Shi, P.-Y., 2021. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. bioRxiv 2021.01.07.425740. https://doi.org/10.1101/2021.01.07.425740

Yin, R., Guest, J.D., Taherzadeh, G., Gowthaman, R., Mittra, I., Quackenbush, J., Pierce, B.G., 2021. Structural and energetic profiling of SARS-CoV-2 antibody recognition and the impact of circulating variants. bioRxiv 2021.03.21.436311. https://doi.org/10.1101/2021.03.21.436311

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Peng, H., Quinlan, B.D., Rangarajan, E.S., Pan, A., Vanderheiden, A., Suthar, M.S., Li, W., Izard, T., Rader, C., Farzan, M., Choe, H., 2020. SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. Nat. Commun. 11, 6013. https://doi.org/10.1038/s41467-020-19808-4

Zhang, W., Davis, B.D., Chen, S.S., Martinez, J.M.S., Plummer, J.T., Vail, E., 2021. Emergence of a novel SARS-CoV-2 strain in Southern California, USA. medRxiv 2021.01.18.21249786. https://doi.org/10.1101/2021.01.18.21249786

Zhou, H., Chen, X., Hu, T., Li, J., Song, H., Liu, Y., Wang, P., Liu, D., Yang, J., Holmes, E.C., Hughes, A.C., Bi, Y., Shi, W., 2020. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. Curr. Biol. 30, 2196-2203.e3. https://doi.org/10.1016/j.cub.2020.05.023

Zhou, H., Ji, J., Chen, X., Bi, Y., Li, J., Wang, Q., Hu, T., Song, H., Zhao, R., Chen, Y., Cui, M., Zhang, Y., Hughes, A.C., Holmes, E.C., Shi, W., 2021. Identification of novel bat coronaviruses sheds light on the evolutionary origins of SARS-CoV-2 and related viruses. Cell. https://doi.org/10.1016/j.cell.2021.06.008

Zhu, X., Mannar, D., Srivastava, S.S., Berezuk, A.M., Demers, J.-P., Saville, J.W., Leopold, K., Li, W., Dimitrov, D.S., Tuttle, K.S., Zhou, S., Chittori, S., Subramaniam, S., 2021. Cryo-electron microscopy structures of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies. PLOS Biol. 19, e3001237. https://doi.org/10.1371/journal.pbio.3001237