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 prior to the emergence of the B.1.1.529 (omicron) lineage. This manuscript describes a single recurrent insertion region (RIR1) in the N-terminal domain of SARS-CoV-2 spike protein, characterized by at least 41 independent acquisitions of 1-8 additional codons between Val213 and Leu216 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), with the quickly spreading omicron and with other VOCs and VOIs warrants further investigation concerning its effects on spike structure and viral infectivity.

Keywords

Covid-19; immune escape; evolution; variant; coronavirus; *Sarbecovirus;* omicron

Abbreviations

ACE2: angiotensin converting enzyme 2

NTD: N-terminal domain

RBD: receptor binding domain

RBM: receptor binding motif

RGYR: radius of gyration

RIR1: recurrent insertion region 1

RMSD: root mean square deviation of backbone beads

RMSF: root mean square fluctuations

38 RDR: recurrent deletion region

39 VOC: variant of concern

40 VOI: variant of interest

41

42 **1. Introduction**

43 Coronaviruses generally accumulate mutations at a much lower rate than other RNA viruses, thanks to the
44 efficient proofreading exonuclease activity exerted by nsp14, in complex the activator protein nsp10
45 (Denison et al., 2011; Ma et al., 2015). As a result, the rate of molecular evolution of SARS-CoV-2 is
46 currently estimated (as of December 5st 2021, based on GISAID data (Shu and McCauley, 2017)), to be close
47 to 25 substitutions/year per genome, i.e. 8.36×10^{-4} substitutions/site/year, which is slightly higher than
48 previous estimates for human endemic coronaviruses (Ren et al., 2015). Consistently with comparative
49 genomics data obtained from other members of the *Sarbecovirus* subgenus, such mutations are not evenly
50 distributed across the genome, but they are disproportionally located in the S gene, which encodes the
51 spike glycoprotein. It is also worth noting that the S gene undergoes frequent recombination events, likely
52 as a result of naturally occurring co-infections in the animal viral reservoirs (Boni et al., 2020), and that
53 these events are also theoretically possible among different SARS-CoV-2 lineages (Ignatieva et al., 2021).
54 The encoded transmembrane protein forms a homotrimer and plays a fundamental role in the interaction
55 between the virus and host cells, promoting viral entry though the interaction with different membrane
56 receptors (Millet et al., 2020). In the case of SARS-CoV-2 and of the closely related SARS-CoV responsible of
57 the 2002-2004 outbreak, such receptor is represented by the angiotensin converting enzyme 2 (ACE2) (Ren
58 et al., 2008; Walls et al., 2020).
59 While most of these mutations have little or no phenotypic impact at all, some may significantly influence
60 viral transmissibility and the ability of the virus to escape host immune response. The causes underpinning
61 such phenotypic effects may either lie in an increased viral shedding, in the alteration of the binding affinity
62 between the spike receptor binding domain (RBD) and the host ACE2 receptor, or in the modification of key
63 antibody epitopes. The most striking example of a non-synonymous mutation which had a dramatic impact
64 on the dynamics of the pandemics is most certainly represented by S:D614G. This mutation, which was not
65 present in the ancestral lineage that caused the Wuhan outbreak, emerged in the very early phases of the
66 pandemics, quickly becoming dominant worldwide (Korber et al., 2020), most likely due to an increased
67 packing of functional spike protein into the virion (Zhang et al., 2020).
68 Even though the mutation rate of the SARS-CoV-2 genome remained relatively stable throughout 2020,
69 growing evidence soon started to point out the presence of shared mutations across multiple independent
70 lineages, suggesting ongoing convergent evolution and possible signatures of host adaptation (van Dorp et
71 al., 2020a). While early investigations failed to identify evidence of increased transmissibility associated
72 with such recurrent mutations (van Dorp et al., 2020b), the nearly contemporary independent emergence
73 of three variants sharing the non-synonymous substitution N501Y in the spike protein started to raise
74 serious concerns about the possible involvement of this mutation in increasing viral infectivity. While the
75 functional role of N501Y still remains to be fully elucidated, structural modeling points towards a possible
76 function in the stabilization of the spike protein in the open conformation, which may increase ACE2
77 binding, especially in combination with other mutations targeting the RBD (Nelson et al., 2021; Teruel et al.,
78 2021; Zhu et al., 2021).
79 B.1.1.7 (the alpha variant, according to WHO labeling), one of the emerging lineages carrying N501Y, spread
80 in southeastern England in early 2020 and quickly became dominant in Europe. Despite being significantly

more transmissible than wild-type genotypes (Davies et al., 2021), alpha 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., 2020; Starr et al., 2021). On the other hand, 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). Following these events, the trends of new infections connected with emerging variants started to be subjected to closer monitoring due to the possible negative impact they might have on massive vaccination campaigns (Madhi et al., 2021; Shen 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 to completely escape the neutralizing activity of several monoclonal antibodies (mAbs) (Starr et al., 2021). 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 (Lopez Bernal et al., 2021). Although delta became dominant worldwide in the second half of 2021, the novel viral lineage B.1.1.529 started to quickly spread in the Gauteng province (South Africa), outcompeting delta. The higher fitness of this lineage has been tentatively linked with a substantial ability to evade immunity from previous infection (Pulliam et al., 2021), which might be consistent with the high number of non-synonymous mutations (32) and indels (three deletions plus one insertion) observed in the S gene compared to the reference SARS-CoV-2 genome. These include a number of previously described RBD mutations associated with the aforementioned VOCs, such as K417N, T478K and N501Y, plus E484A. Based on early epidemiological data, on the growing number of imported cases reported abroad and on the expected immune evasion properties of this variant, on December 1st WHO included B.1.1.529 in the list of VOCs under the “omicron” designation.

Several VOCs and VOIs (including alpha, beta, delta and omicron) carry spike deletions in the NTD. Such deletions were previously shown to often occur in distinct NTD sites, named Recurrent Deletion Regions (RDR), 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, prior to the emergence of omicron, which carries a three amino

127 acids-long insertion (S:ins214EPE) in the NTD of the spike protein, very little attention had been directed
128 towards insertions. Nevertheless, such events are known to have played a fundamental role in the past
129 evolution of SARS-CoV-2 spike protein by allowing, among the other things, the acquisition of a furin-like
130 cleavage site, which is an uncommon feature in bat coronaviruses. This short motif, which is thought to be
131 a key pathogenicity determinant (Johnson et al., 2021), is indeed completely absent in the closely related
132 *Sarbecovirus* RaTG13 (Ge et al., 2016) and only partly present in the recently described RmYN02 (Zhou et
133 al., 2020) and RacCS203 (Wacharapluesadee et al., 2021).
134 The present work reports the independent occurrence of at least 41 distinct insertion events at the very
135 same NTD site, located between Val213 and Leu216, which will be hereafter referred to as Recurrent
136 Insertion Region 1 (RIR1). The transient international spread of the RIR1 insertion-carrying lineages A.2.5
137 and B.1.214.2, the presence of S:ins214EPE in the VOC omicron and the identification of several insertions
138 at this site in the delta lineage point out that more attention should be put towards the functional
139 characterization of these codon acquisitions in the near future.
140

141 **2. Materials and methods**

142

143 **2.1. Sequence data analysis**

144

145 The global frequency of insertion and deletion mutations mapped on the SARS-CoV-2 S gene was retrieved,
146 based on GISAID data (Shu and McCauley, 2017), from https://mendel.bii.a-star.edu.sg/ (last accessed on
147 December 5th, 2021; credit to Raphael Tze Chuen Lee). Disruptive insertion and deletion mutations (i.e.
148 those that interrupted the open reading frame of the S gene) and insertions carrying undetermined amino
149 acids were discarded. Genomes carrying insertions at any position between codons 213 and 216 were
150 grouped based on the inserted nucleotide sequence. Each group was assigned a code based on progressive
151 Roman numerals, following their chronological order of identification; a lowercase letter was added to
152 design variants of the same insertion that determined a change in the encoded amino acid sequence. The
153 nucleotide sequences of representative entries for each of the identified insertions were aligned with the
154 Wuhan-Hu-1 isolate SARS-CoV-2 reference sequence (GenBank ID: NC_045512.2) using MUSCLE (Edgar,
155 2004) in the MEGA X environment (Kumar et al., 2018), initially preserving codon boundaries. The multiple
156 sequence alignment was then manually refined to reflect the most probable location of the insertion within
157 each codon and each event was consequently classified as a phase 0, phase I or phase II insertion.
158 All SARS-CoV-2 genome data used for phylogenetic inference in this study were retrieved from GISAID (last
159 access date December 5th, 2021) (Shu and McCauley, 2017). In detail, all available sequenced genomes
160 belonging to the lineage A.2.5, to the related sublineages A.2.5.1, A.2.5.2 and A.2.5.3, and to the sister
161 lineage A.2.4 were downloaded, along with associated metadata. While all available GISAID entries were
162 considered for reporting observation frequencies, only high quality genomes (i.e. those listed as "complete"
163 and "high coverage") associated with a sampling date were taken into account for further analysis.
164 Genomes containing long stretches of Ns (i.e. comprising more than 25 consecutive undetermined
165 nucleotides) were discarded. The reference isolate Wuhan-Hu-1 was also included for tree rooting
166 purposes. Note that several genome sequences from Panama with sampling date anterior to November
167 2021 were discarded due to the unreliability of associated metadata (i.e. the sampling dates appeared to
168 be inconsistent with the very small genetic distances with recent isolates belonging to the same lineage).
169 Overall, the A.2.5-focused datasets included 1,283 sequences.
170 SARS-CoV-2 genomes were analyzed with the nextstrain *augur* pipeline
171 (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 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. The phylogenetic tree 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.

A root-to-tip genetic distance analysis was performed by plotting the sampling dates against the total number of nucleotide substitutions (excluding insertions and deletions) observed in genomes belonging to the A.2.5 lineage and related sublineages. These were calculated with MEGA X (Kumar et al., 2018), compared with the reference genotype Wuhan-Hu-1. The global average genome-wide mutation rate of SARS-CoV-2, roughly equivalent to 25 substitutions per year, was retrieved from GISAID (as of December 5th, 2021).

2.2. 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 structure of the spike protein of A.2.5 with Swiss Model (Waterhouse et al., 2018), using the EPI_ISL_1502836 GISAID entry as a reference. The protein structure was converted to a coarse-grained Martini representation 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) 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).

3. Results and Discussion

3.1. 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 December 5th, 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 0.3% of the total. Prior to the emergence of omicron, the impact of insertions in the S gene on viral evolution had been very limited compared with deletions, which are found in several widespread VOCs and VOIs, such as alpha, beta and delta. Overall, the frequency of observation of spike deletions was more than 500 folds higher than spike insertions, even though this ratio is expected to significantly change in the future, along with the spread of omicron. As previously reported by other authors, most deletions occur in specific sites of the N-terminal domain, including the four previously identified Recurrent Deletion Regions (RDR) 1, 2, 3 and 4 (**Figure 1**) (McCarthy et al., 2021) and the deletion which characterizes the delta variant, occurring at positions 157/165. 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 (nearly 3,000 genomes) identified at a specific site, located between codons 213 and 216 (**Figure 1**). However, this figure might be an underestimate due to the frequent use of reference-based insertion-unaware algorithms for SARS-CoV-2 genome assembly, especially during the early phases of the pandemics. 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 at several other spike sites, RIR1 was the only one where multiple insertions have independently occurred in different lineages. The only spike insertions site displaying a higher absolute number of sequenced SARS-CoV-2 genomes (as of December 5th, 2021) is S:ins145T, found in the VOI mu (Laiton-Donato et al., 2021) (**Figure 1**).

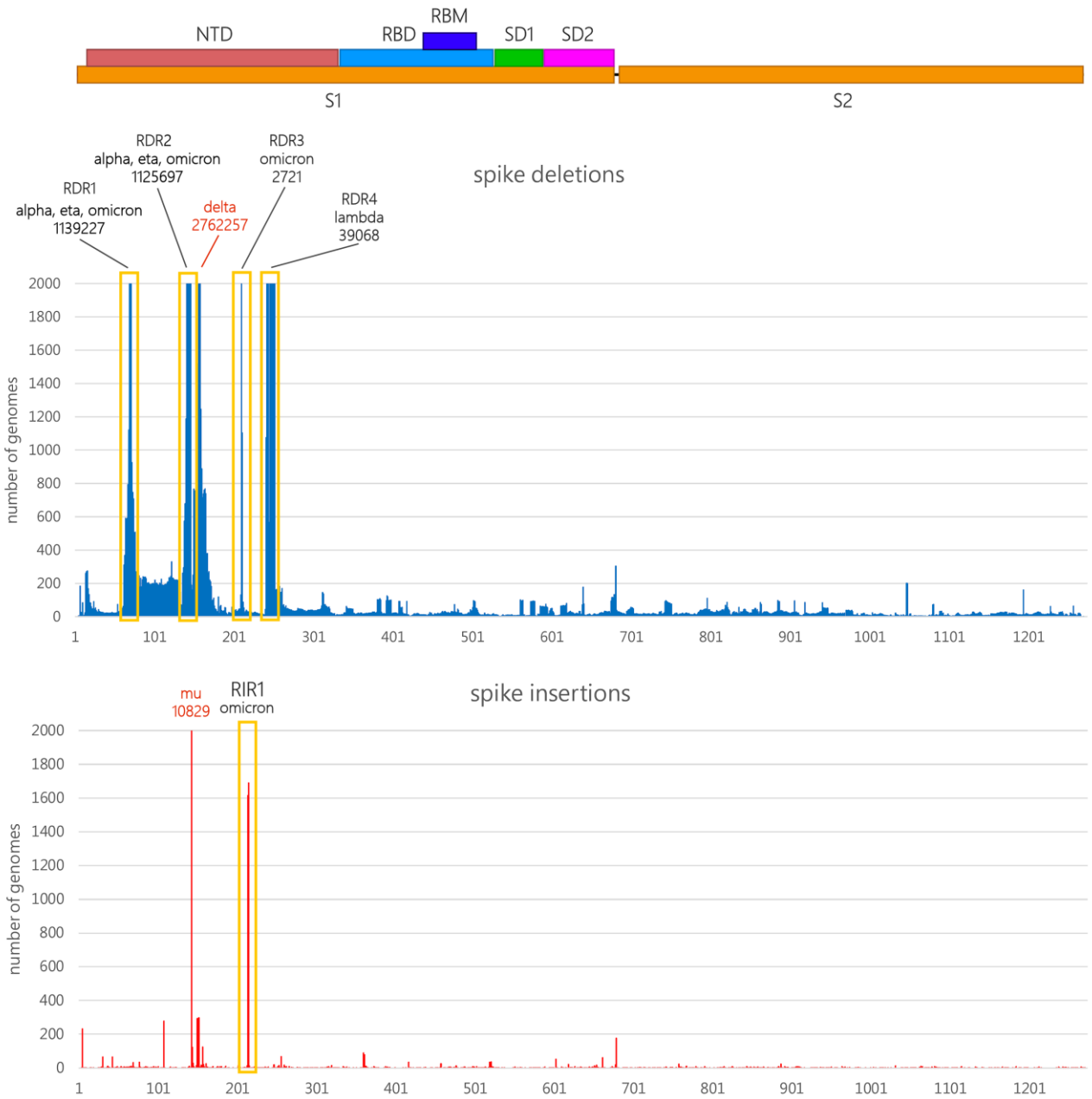


Figure 1. Schematic representation of the SARS-CoV-2 protein, with indication of the two functional S1 and S2 subunits, which are separated by a furin-like proteolytic cleavage site, the N-terminal domain (NTD), the receptor binding domain (RBD) and receptor binding motif (RBM), the SD1 and SD2 subdomains. The absolute number of observed deletion mutations along the S-gene are reported (https://mendel.bii.a-star.edu.sg/ was last accessed on December 5th, 2021). Bars were truncated at 2,000 observed genomes; in such cases, the absolute number of observations is reported above the truncated bars, together with the main VOCs and VOIs associated with each indel. The position of RDR1-RDR4 from a previous study (McCarthy et al., 2021), as well as the deletion 157/158 characterizing the delta variant and the ins145T insertion characterizing the mu variant, are reported.

3.2. RIR1 insertions independently emerged in multiple viral lineages

As of December 5th, 2021 RIR1 insertions could be documented as the result of at least 41 independent events that occurred in different branches of the SARS-CoV-2 phylogenetic tree, which strongly suggests convergent evolution. Even though the length of the insertion spanned from one to eight codons (**Figure 2**),

253 the overwhelming majority of the genomes with RIR1 insertions (99% of the total) only included three
254 codons (**Table 1**). While most RIR1 insertions were associated with very small local clusters that not lead to
255 further spread, a few were found in viral lineages with widespread community transmission. These include
256 the lineages B.1.1.529/omicron (insertion XXXIX), A.2.5 (insertion III), B.1.214.2 (insertion IV) and also
257 B.1.639 (insertion IX), which was never associated with a high number of infection but displayed continuous
258 circulation for several months (**Table I**). While omicron is currently the subject of intense scrutiny by the
259 scientific community, A.2.5 and B.1.214.2 were never included in the list of VOIs, despite reaching high
260 prevalence in some geographical regions. While the former lineage will be discussed in detail as a case
261 study in the following sections, it is worth briefly mentioning here the transient spread of B.1.214.2 in early
262 2021. Indeed, this lineage was likely imported from central Africa to Europe in late 2020. Here, between
263 March and April 2021, it accounted for a non-negligible fraction of the covid-19 cases recorded in Belgium
264 and Switzerland. Following an initial spread, which led to over 1,000 documented infections worldwide, the
265 frequency of observation of B.1.214.2 started to drop in parallel with the rise of alpha, and further declined
266 when delta became dominant. This lineage has not been detected since early July 2021 and can be thus
267 provisionally considered as extinct. Insertion IV, which results in the addition of the TDR tripeptide between
268 R214 and D215, was associated with the presence of two other spike non-synonymous mutations located
269 on the RBD (i.e. Q414K and N450K). These have been previously linked with a moderately increase in RBD
270 stability (Teng et al., 2021) and immune escape both towards a few mAbs and towards convalescent sera
271 (Liu et al., 2021), respectively. Due to the lack of functional data, it is presently unknown whether insertion
272 IV and the other aforementioned mutations endowed this lineage with improved transmissibility or with
273 increased potential for reinfection.
274 Several other insertion events at RIR1 occurred in lineages identified as VOCs or VOIs by WHO, CDC, ECDC
275 or PHE, including some that have been recently de-escalated to the status of variants under monitoring. In
276 detail, insertion V was found in eight viral genomes belonging to the gamma lineage, sequenced in different
277 Brazilian states and Guyana between December 2020 and April 2021, indicating the presence of community
278 transmission in the region. As reported in a previous work (Resende et al., 2021), these genomes belong to
279 a monophyletic P.1-like clade that appears to be basal to P.1. The highly transmissible alpha lineage, which
280 became dominant in Europe and quickly spread worldwide in early 2021 (Volz et al., 2021), before the rise
281 of delta, was associated with at least five independent insertions at RIR1 (insertion XI, XII, XIII, XV and XX)
282 between February and May 2021 (**Table 1**, **Figure 2**). A single RIR1 insertion (XXX) was recorded in August
283 2021 in the lineage B.1.525 (eta) in the United Kingdom (Ozer et al., 2021) and two sequence genomes
284 characterized by the presence of insertion VII belonging to B.1.429 (epsilon) (Zhang et al., 2021) were
285 sequenced in California in January 2021. Several recently identified insertions at RIR1 are associated with
286 delta (i.e. insertion XXIII, XXV, XXVI, XXVII, XXXII, XXXIII, XXXIV, XXXV, XXXVI, XXVII, XXXVIII, XL and XLI), but
287 none of these have led to significant community spread to date, even though a couple of them were linked
288 to small clusters in England (**Table 1**, **Figure 2**).
289 One of the most recent RIR1 insertions (insertion XXXIX) was associated with the emergence of the
290 B.1.1.529/omicron lineage, first detected on November 8th, 2021, but possibly spreading undetected in
291 Southern Africa since early October. While the number of omicron cases confirmed by genome sequencing
292 as of December 5th, 2021 is still relatively low (see **Table 1**), the rapid spread of this variant worldwide
293 strongly suggest that S:ins215EPE will become the most frequently observed RIR1 insertion in the matter of
294 a few days.
295 Albeit not directly linked with variants designated as VOCs or VOIs, other RIR1 insertions were associated
296 with the presence of immunologically relevant spike mutations. This is the case of the previously
297 mentioned insertion IX (lineage B.1.639), which is characterized by the contemporary presence of E484K,

T478K and by the deletions Δ69/Δ70 (found in RDR1) and Δ144 (found in RDR2), which are shared by several VOCs and VOIs. Curiously, like the omicron insertion XXXIX, insertions XIV, XIX, XXI, XXII and XXXI also targeted viral genomes carrying non-synonymous spike mutations at E484 and deletions at RDR1, suggesting a possible role of RIR1 insertions in compensating otherwise slightly deleterious mutations, like previously hypothesized for RDR1 (Meng et al., 2021).

Taking into account the limited efforts carried out by several countries in genomic surveillance throughout 2020 and 2021, the insertions reported in **Table 1** and **Figure 2** may just represent a fraction of those that emerged at RIR1 during the course of the pandemics. Although it was possible to unambiguously ascertain the exact placement of just 27 out of the 41 RIR1 insertions (see **Table 1**), most of them (17 out of 27 cases, i.e. 63%) were in-frame, occurring at phase 0 between codons 214 and 215, with no effect on neighboring codons. However, others were out-of-frame, occurring either at phase I (i.e. between the first and the second nucleotide of a codon) or at phase II (i.e. between the second and the third nucleotide of codon) (**Figure 2**). In detail, three insertions were observed at phase II within codon 213, one at phase II within codon 214, three and two at phase I and II, respectively, within codon 215, and a single one at phase I within codon 216. In such cases, the placement of the insertion often determined a non-synonymous mutation of the residues flanking RIR1 either at the N- or at the C-terminal side (**Table 1**). It is also worth noting that insertion XXXIX (B.1.1.529/omicron) was associated with a three nucleotides-long, out-of-frame proximal deletion, which affects codons 211 and 212, resulting in the deletion of a single amino acid. Similar deletions were uncommon in other lineages carrying RIR1 insertions, as they have been previously observed in a single other case, i.e. insertion XXI, which displays a Δ210 deletion.

Although the origins of the 41 RIR1 insertions was not investigated in the present study, other authors have previously suggested that they may be result either from the incorporation of distal regions of the SARS-CoV-2 genome itself, of host mRNAs (Peacock et al., 2021), or even of the genomic sequence of endemic coronaviruses co-infecting the host (Venkatakrishnan et al., 2021). These events would be most likely explained by a still poorly understood copy-choice recombination processes occurring during viral genome replication (Chrisman et al., 2021; Garushyants et al., 2021). Nevertheless, we caution that the short length of RIR1 insertions (usually nine to twelve nucleotides) is in most cases not sufficient to unequivocally establish the origins of the inserted nucleotide sequence, since several randomly occurring identical sequence matches are expected to be found in a broad range of living organisms. On the other hand, when RIR1 insertions are relatively long, the robustness of such inferences might be significantly higher (Peacock et al., 2021).

Designation	Insertion	Insertion type	Lineage	GISAID entries	Other spike mutations	Earliest detection	Latest detection*
I	214:AKKN	ambiguous (codon 215 phase0/I)	B	1	none	Mar 5th, 2020	Mar 5th, 2020
II	214:KLGP	in-frame (codon 215 phase 0)	B.1.177	1	E154K, A222V, D614G	Nov 13th, 2020	Nov 13th, 2020
III	214:AAG	out-of-frame (codon 215 phase I)	A.2.5	1691	L141del, G142del, V143del, D215Y, L452R, D614G	Nov 20th, 2020	Nov 5th, 2021
IIIa	214:AVG	out-of-frame (codon 215 phase I)	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	Feb 18th, 2021	Feb 18th, 2021
IIIb	214:ASG	out-of-frame (codon 215 phase I)	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	Jan 21st, 2021	Jan 21st, 2021
IIIc	214:AAV	out-of-frame (codon 215 phase I)	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	May 17th, 2021	May 17th, 2021
IIId	214:DGR	out-of-frame (codon 215 phase I)	A.2.5	1	L141del, G142del, V143del, D215L, L452R, D614G	Jan 19th, 2021	Jan 19th, 2021
IIIe	214:AAGAAG	out-of-frame (codon 215 phase I)	A.2.5	1	L141del, G142del, V143del, D215Y, L452R, D614G	May 6th, 2021	May 6th, 2021
IV	214:TDR	in-frame (codon 215 phase 0)	B.1.214.2	1224	Q414K, N450K, D614G, T716I	Nov 22nd, 2020	Jun 28th, 2021
IVa	214:IDR	in-frame (codon 215 phase 0)	B.1.214.2	1	Q414K, N450K, D614G, T716I	Apr 3rd, 2021	Apr 3rd, 2021
IVb	214:QID	in-frame (codon 215 phase 0)	B.1.214.2	3	Q414K, N450K, D614G, T716I	Apr 12th, 2021	May 1st, 2021
V	214:ANRN	ambiguous (codon 215 phase0/I)	P.1	9	L18F, P26S, D138Y, K417T, E484K, N501Y, D614G, D1139H, V1176F	Dec 23rd, 2020	Apr 5th, 2021
VI	214:KRI	in-frame (codon 215 phase 0)	B	4	V367F, E990A	Dec 28th, 2020	Mar 15th, 2021
VII	214:AQER	ambiguous (codon 215 phase0/I)	B.1.429	2	S13I, P26S, S98F, W152C, L452R, D614G, T1027I	Jan 15th, 2021	Jan 18th, 2021
VIII	214:DLA	ambiguous (codon 215 phase0/i/II, codon 216 phase 0/I/II)	B.1.2	5	D614G	Jan 17th, 2021	Feb 2nd, 2021
IX	214:QAS	in-frame (codon 215 phase 0)	B.1.639	45	H69del, V70del, Y144del, M153T, T478K, E484K, D614G, T859N, D936Y	Jan 19th, 2021	Nov 4th, 2021
X	214:DRS	out-of-frame (codon 215 phaseI/II)	B.1	1	D215N, V382L, D614G, M1237I	Feb 1st, 2021	Feb 1st, 2021
XI	214:KFH	in-frame (codon 215 phase 0)	B.1.1.7	3	H69del, V70del, Y144del, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Feb 12th, 2021	Feb 22nd, 2021
XII	214:KAFKQ	in-frame (codon 215 phase 0)	B.1.1.7	2	H69del, V70del, Y144del, A262S, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Feb 25th, 2021	Feb 25th, 2021
XIII	214:ER	out-of-frame (codon 215 phase II)	B.1.1.7	1	H69del, V70del, Y144del, D215G, N501Y, A570D, D614G, P681H, I712V, T716I, S982A, D1118H	Mar 11th, 2021	Mar 11th, 2021
XIV	214:NWAHW	in-frame (codon 215 phase 0)	B.1.547	1	T19I, T95I, H69del, V70del, D614G, E484A, A879T, T1027I	Mar 22nd, 2021	Mar 22nd, 2021
XV	214:APR	ambiguous (codon 215 phase0/I)	B.1.1.7	1	H69del, V70del, Y144del, A262S, N501Y, A570D, D614G, P681H, T716I, S982A, D1118H	Mar 31st, 2021	Mar 31st, 2021
XVI	213:KAR	out-of-frame (codon 213 phase II)	B.1.177	1	A222V, A262S, P272L, D614G, P681H, M1229I	Apr 2nd, 2021	Apr 2nd, 2021
XVII	214:KLRS	ambiguous (codon 215 phase II, codon 216 phase 0)	A.28	6	T76I, D215S, N501T, F592S, H655Y	Apr 23rd, 2021	Jun 22nd, 2021
XVIII	214:KGE	in-frame (codon 215 phase 0)	B.1.1.519	1	T732A, T478K, D614G, P681H	Apr 24th, 2021	Apr 24th, 2021
XIX	214:W	in-frame (codon 215 phase 0)	B.1		T19I, F140del, P139del, L141del, G142del, V143del, Y144del, Y145del, I210del, L242del, A243del, L244del, T470N, S494P, D614G, H655Y, T859N	May 14th, 2021	May 14th, 2021
XX	216:QMR	out-of-frame (codon 216 phase I)	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
XXI	214:KLLIRGD	in-frame (codon 215 phase 0)	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
XXII	213:KRLR	out-of-frame (codon 213 phase II)	B.1.1	1	R214Q, D614G, E484D	Jun 6th, 2021	Jun 6th, 2021
XXIII	214:EAR	ambiguous (codon 215 phase I/II/III)	B.1.617.2	1	T19R, N137K, G142D, E156G, F157del, R158del, L452R, T478K, E484Q, D614G, P681R, D950N	Jun 8th, 2021	Jun 8th, 2021
XXIV	214:QCEE	in-frame (codon 215 phase 0)	B.1.247	1	P209S, A222V, T572I	Jul 15th, 2021	Jul 15th, 2021
XXV	214:ELCD	ambiguous (codon 215 phase I/II/III)	AY.12	1	T19R, T95I, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Jul 17th, 2021	Jul 17th, 2022
XXVI	214:NFGGG	in-frame (codon 215 phase 0)	AY.4	2	T19R, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Jul 22nd, 2021	Jul 22nd, 2021
XXVII	214:NES	in-frame (codon 215 phase 0)	AY.16	1	T19R, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Jul 30th, 2021	Jul 30th, 2021
XXVIII	214:VPWI	ambiguous (codon 215 phase 0/I)	AY.4	9	T19R, T95I, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R,V622F, D950N	Aug 4th, 2020	Aug 21st, 2021
XXIX	214:HSG	in-frame (codon 215 phase 0)	AY.4	7	T19R, T95I, D138H, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Aug 4th, 2021	Aug 25st, 2021
XXX	214:SYCTKSKC	in-frame (codon 215 phase 0)	B.1.525	1	A67V, H69del, V70del, Y144del, E484K, D614G, A653V, N679del, Q677H, F888L	Aug 5th, 2021	Aug 5th, 2021
XXXI	214:ESH	ambiguous (codon 215 phase 0/I/II)	B.1.240	1	C15F, L141del, G142del, V143del, Y144del, L242del, A243del, L244del, G446V, E484A, D614G, A688V, V1176F	Aug 6th, 2021	Aug 6th, 2021
XXXII	214:NFG	in-frame (codon 215 phase 0)	AY.25	3	T19R, S112L, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Aug 20th, 2021	Sep 14th, 2021
XXXIII	214:DQAF	out-of-frame (codon 214 phase II)	B.1.617.2	1	T19R, G142D, E156G, F157del, R158del, A222V, V289I, L452R, T478K, D614G, P681R, D950N	Aug 21st, 2021	Aug 21st, 2021
XXXIV	214:EGAE	ambiguous (codon 215 phase 0/I/II)	AY.4	2	T19R, T95I, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Sep 2nd, 2021	Sep 2nd, 2021
XXXV	214:RLR	in-frame (codon 215 phase 0)	AY.3	2	T19R, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Sep 20th, 2021	Sep 20th, 2021
XXXVI	214:GLKG	out-of-frame (codon 215 phase I)	AY.101	1	T19R, T95I, G142D, E156G, F157del, R158del, L452R, T478K,D614G, P681R, D950N	Oct 1st, 2021	Oct 1st, 2021
XXXVII	215:EAR	out-of-frame (codon 215 phase II)	AY.4	1	T19R, T95I, E156G, F157del, R158del, D215N, L452R, T478K, D614G, P681R, D950N	Oct 3rd, 2021	Oct 3rd, 2021
XXXVIII	214:VRR	out-of-frame (codon 215 phase I)	AY.113	1	T19R, G142D, E156G, F157del, R158del, D215H, L452R, D614G, P681R, D950N	Oct 26th, 2021	Oct 26th, 2021
XXXIX	214:EPE	ambiguous (codon 215 phase 0/I/II)	B.1.1.520	533	A67V, H69del, V70del, T95I, V143del, Y144del, Y145del, G142D, N211del, L212I, G339D, S371L, S373P, S375F, K417N, G446S, S477N, T478K, Q493R, G496S, Q498R, N501Y, Y505H,T547K, D614G, H655Y, N764K, N679K, D796Y, E484A, P681H, N856K, Q954H, N969K, L981F	Nov 8th, 2021	Dec 2nd, 2021
XL	215:KGD	ambiguos (codon 215 phase II, codon 216 phase 0)	AY.4.2	1	T19R,T95I,Y145H,E156G,F157del,R158del,G142D,A222V,L452R,T478K,D614G,P681R,D950N	Nov 18th, 2021	Nov 18th, 2021
XLI	213:RWR	out-of-frame (codon 213 phase II)	AY.4	2	T19R, T95I, G142D, E156G, F157del, R158del, L452R, T478K, D614G, P681R, D950N	Nov 19th, 2021	Nov 22nd, 2021

*as of December 5th, 2021

Table 1. Summary of the 41 independent RIR1 insertions found in the SARS-CoV-2 genome, ordered by the earliest date of detection, as of December 5th, 2021.

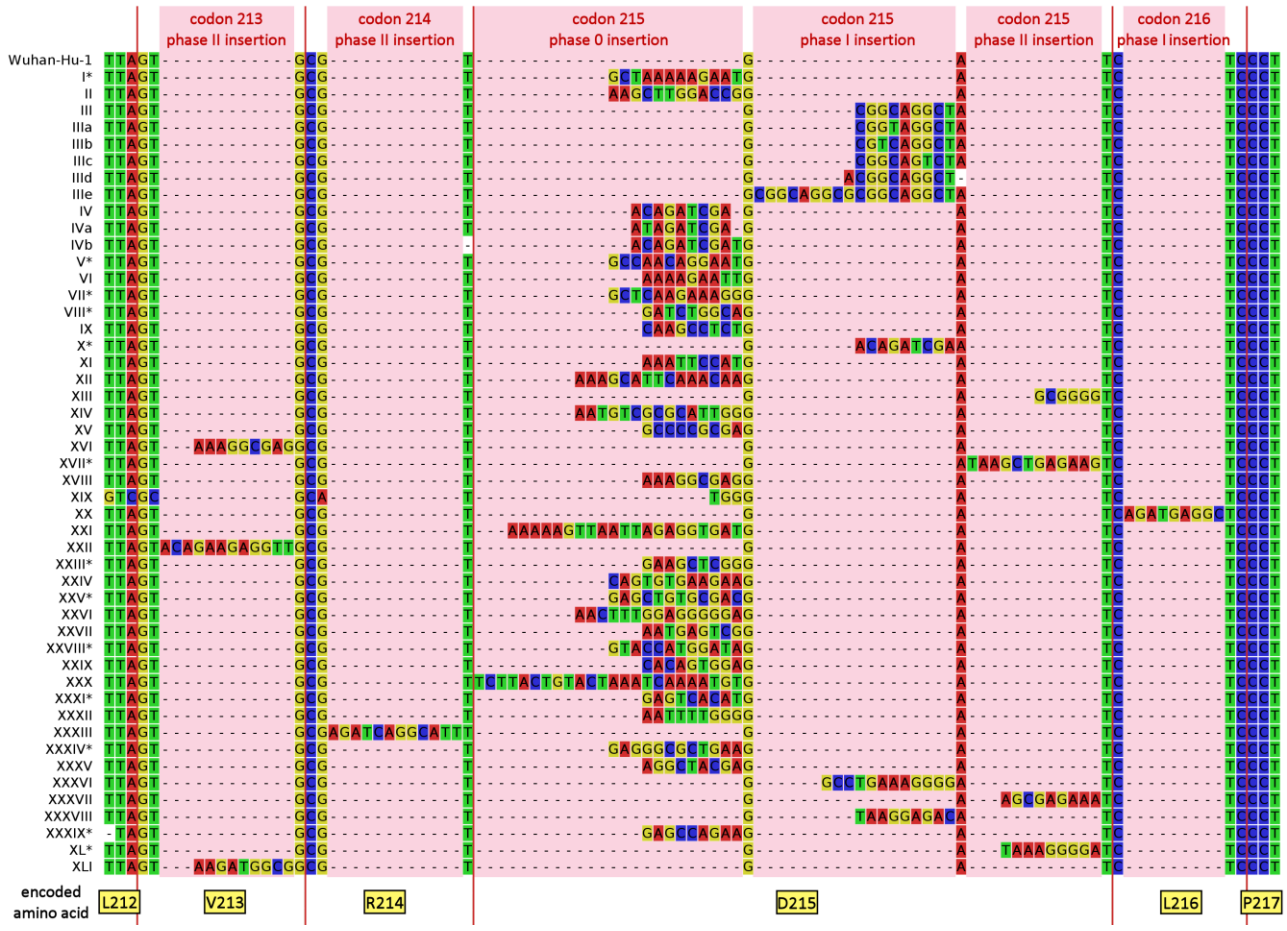


Figure 2. Multiple sequence alignment of the nucleotide sequences of the SARS-CoV-2 S gene of the viral lineages characterized by an insertion at RIR1, compared with the reference sequence Wuhan Hu-1. The multiple sequence alignment only displays a small portion of the S gene and of the encoded spike 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. Please note that the exact position of all insertion could not be unambiguously detected in all cases; those with ambiguous placement are marked with an asterisk.

3.3. Mutational pattern of A.2.5 lineage

As mentioned above, the only two lineages carrying 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 within codon 215, 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 lineages currently or previously designated as VOCs and VOIs, L452R is also found

in delta, kappa and epsilon. Like the overwhelming majority of the variants circulating in 2021, A.2.5 is characterized by the presence of the prevalent mutation D614G. Although no other spike mutations are widespread in A.2.5, the A.2.5.3 sublineage acquired S477N, also found in omicron and known to strengthen the binding with the ACE2 receptor (Singh et al., 2021). Overall, this mutation is associated with ~2% of all A.2.5 genomes (**Figure 3A**). Other relevant spike non-synonymous mutations, known to significantly alter either ACE2 binding or antibody recognition, were only seldom detected: K417T, N501Y and E484K (which are the hallmark spike mutations of gamma) were simultaneously found in a single genome (EPI_ISL_2305075, see **Figure 3A**) sequenced in Texas in May 2021. E484K was found in three additional cases (two in the United States, one in Canada) in April 2021, and N501Y was detected in nine additional cases (eight in the United States, one in Canada) between March and May 2021. Interestingly, in one such cases N501Y was paired with E484Q, which is found in kappa and determines reduced antibody sensitivity, even though not synergistically with L452R (Ferreira et al., 2021). The acquisition of the mutation P681H, known to increase the efficiency of the furin-like cleavage site was documented in 8 cases, 6 of which also displayed N501Y. Such insertions occurred independently in different branches of the A.2.5 evolutionary tree, indicating convergent evolution (see **Figure 3A**). Other lineage-defining non-synonymous mutations of A.2.5 are placed in other genomic locations. These included K1657E, F3071Y, T3255I (shared with delta and mu) and H3580Q in ORF1a; P1000L in ORF1b; S74F and G196V in ORF3a; S197L and M234I (shared with iota) in N (see **Figure 3B**). The functional consequences of these point mutations are presently unknown.

Root-to-tip genetic distance analysis revealed that the overall nucleotide substitution rate observed in the A.2.5 lineage (and related sublineages) was significantly lower than the average substitution rate computed for SARS-CoV-2 (based on GISAID data, last retrieved on December 5th, 2021), as evidenced by the markedly different slope of the regression line (see **Figure 3C**). This was consistent with a substitution rate equal to 4.00×10^{-4} substitutions/site/year, i.e. roughly 12 substitutions/genome/year. Nevertheless, the A.2.5 SARS-CoV-2 genomes detected in the earliest phases of the spread of this lineage (i.e. December 2020) were linked with a number of substitutions significantly higher than the average number of substitutions found in the same period in other SARS-CoV-2 lineages (i.e. ~33 vs ~25).

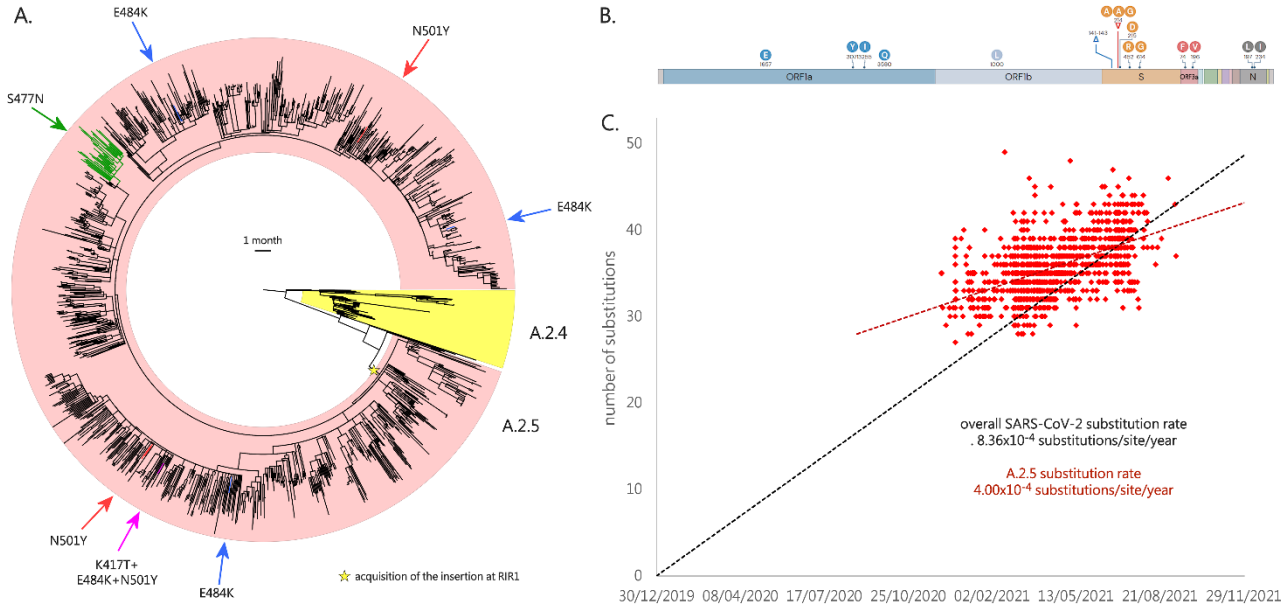
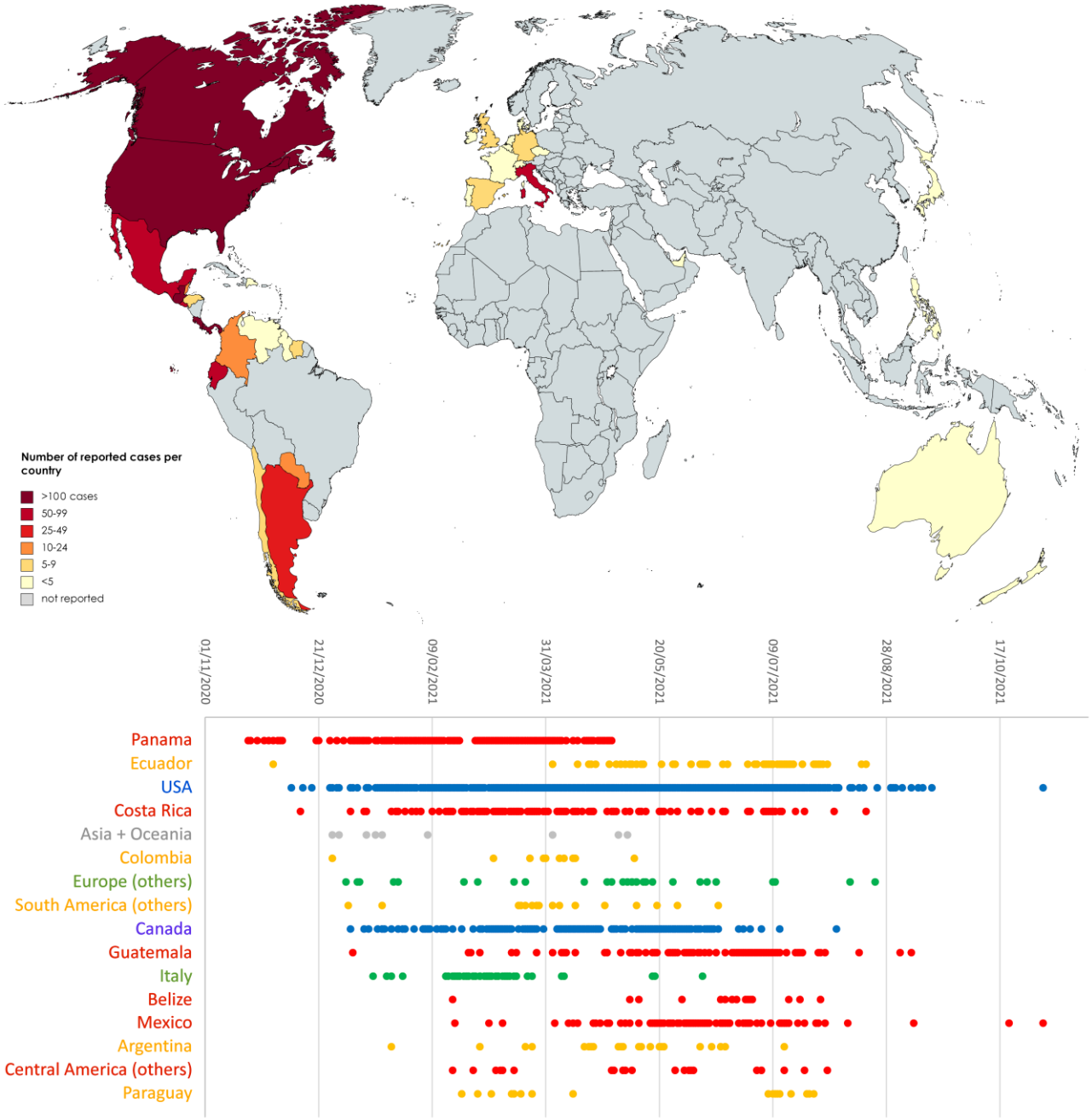


Figure 3. Panel A: circular time tree exemplifying the phylogeny of the A.2.5 lineage related sublineages. Only high quality, complete genomes have been included. The Wuhan-Hu-1 strain was used to root the tree; the sister lineage A.2.4 is also indicated. The acquisition of relevant spike mutations placed in the

389 receptor binding domain (i.e. S477N, K417T, E484K and N501Y) is marked with arrows. Please note that the
390 monophyletic clade linked with the acquisition of S477N corresponds to the A.2.5.3 sublineage. Panel B:
391 key mutations associated with the A.2. lineages. Genes associated with mutations (compared with the
392 reference strain Wuhan-Hu-1) are indicated; only mutations detected in > 50% of the genomes belonging to
393 this lineage and associated sublineages are shown. Modified from https://outbreak.info/. Panel C: root-to-
394 tip genetic distance (number of nucleotide substitutions) of the genomes belonging to the A.2.5 lineage and
395 related sublineages, compared with the reference genome Wuhan-Hu-1. The black dashed line represents
396 the average rate of mutation of all SARS-CoV-2 sequenced genomes, according to GISAID (i.e. 25
397 substitutions per genome per year, as of December 5th, 2021). The red dashed line represent the rate of
398 mutation computed for A.2.5. Note that insertions and deletions were excluded from this calculation.
399
400 **3.4. Emergence and international spread of A.2.5**
401 A.2.5 belongs to one of the very few surviving children lineages of the ancestral lineage A, which, after
402 several months of limited global spread, has led to a few major clusters of infections in 2021, such as the
403 one which involved A.23.1 in Uganda (Bugembe et al., 2021). A.2.5 stems from A.2.4, the dominant lineage
404 in the Panama pandemics during the first half of 2020 (Franco et al., 2020). The first documented cases can
405 be traced back to late November 2020, all within a 100 km^2 area around the capital city Panamá. However,
406 the precise timing of the emergence of A.2.5, along with the acquisition of insertion III at RIR1 and of the
407 other associated mutations described in the previous section, is presently unclear due to the insufficient
408 molecular surveillance carried out in Central America. To date, less than 1,300 out of nearly 480K covid-19
409 cases reported in Panama have been selected for viral characterization by sequencing i.e. less than 0.3% of
410 the total, far below of the threshold that would be sufficient to track emerging variants (Vavrek et al.,
411 2021). The presence of a number of genomes sampled in El Salvador and Guatemala, two countries where
412 genomic surveillance has been virtually non-existing in 2020, in the earliest-branching clade belonging to
413 A.2.5 (**Figure 3A**), leaves the precise geographical origins of this lineage unclear.
414 Nevertheless, A.2.5 undoubtedly underwent expansion in Panama between December 2020 and February
415 2021, as revealed by the increase in estimated prevalence from ~60 to ~95%. Interestingly, A.2.5 has been
416 linked with clinically documented reinfections in individuals previously infected by the A.2.4 lineage, which
417 is consistent with the presence of the constellation of non-synonymous spike mutations reported in the
418 previous section, some of which may have immune escape properties (Díaz et al., 2021). The A.2.5 lineage
419 likely spread very early also in the neighboring countries: while investigations carried out in August 2020
420 failed to identify A.2.5 in Costa Rica (Molina-Mora et al., 2021), the prevalence of this lineage in the country
421 reached 30% between March and June 2021, with the establishment of large clusters of community
422 transmission (**Figure 4**). A.2.5 may have undergone a similar spread in other countries in central America,
423 including Belize, Honduras, El Salvador, Guatemala and Mexico, where multiple cases have been detected,
424 starting from the spring of 2021 (**Figure 4**).
425 The remarkable spread of SARS-CoV-2 in Central America was connected with a significant number of
426 exported cases, which have sometimes led to clusters of infection abroad. The first evidence of the
427 detection of A.2.5 in southern America dates to December 1st 2020, in Ecuador. In this country, the
428 acquisition of the spike mutation S477N, mentioned in the previous section, later led to the establishment
429 of the A.2.5.3 sublineage (**Figures 3 and 4**). Reports in other Latin American countries remain sporadic, but
430 it is worth noting that A.2.5 genomes have been so far sequenced in Argentina, Suriname, Guyana,
431 Grenada, Dominican Republic, Sint Marteen, Cayman Islands, Chile, Colombia, Venezuela and Paraguay
432 (**Figure 4**). The earliest cases exported in other continents were reported with similar timing in UAE
433 (December 27th, 2020), Philippines (December 30th, 2020) and Australia (January 11th, 2021), which is
434 consistent with the period with the highest incidence of covid-19 infections documented in Panama.

Cases linked with A.2.5 in Europe were identified in Luxembourg, Portugal, Germany, Italy United Kingdom, Czech Republic, France, Belgium, Ireland, Switzerland, Denmark, Netherlands and Spain. 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 (**Figure 4**). Similarly, imported cases have most certainly led to local cluster of infections in different areas of the United States and Canada, starting from late 2020 (**Figure 4**). Nevertheless, the prevalence of A.2.5 in Northern America never exceeded 0.5%). No SARS-CoV-2 infections linked with A.2.5 have been identified to date in the African continent.

The global frequency of observation of A.2.5 and related sublineages underwent a rapid decline in the second half of 2021, in parallel with the global spread of delta. As of December 5th, 2021, just five genomes belonging to this lineage have been sequenced over the past three months (**Figure 4**). Further molecular surveillance are expected to reveal whether A.2.5 will disappear, in a similar fashion to what occurred for B.1.214.2 during the summer of 2021.



449
450 **Figure 4**. Upper panel: global spread of A.2.5 and related sublineages. Lower panel: detailed timing of the
451 detection of sequenced genomes belonging to A.2.5 and related sublineages in different countries. Only
452 countries with >= 10 unique days of detection are reported, whereas the others were collapsed in
453 geographic macroareas (i.e. Asia + Oceania, Europe, South America and Central America). The reported
454 dates refer to the dates of sampling reported in GISAID.
455
456 **3.5. Impact of RIR1 insertions on the structure of the spike glycoprotein**
457 RIR1 is located in a loop which connects the spike NTD β strands 15 and 16, a region which, unlike RDR1, 2,
458 3 and 4, does not show any overlap with any known major NTD antigenic sites (Cerutti et al., 2021;
459 McCallum et al., 2021). Hence, the involvement of the insertions reported in this manuscript in antibody
460 escape is unlikely, even though the possibility that this modification may lead to paired structural
461 alterations at distantly related sites, leading to a reduced surface accessibility of canonical antibody
462 epitopes cannot be ruled out. Moreover, the possibility that RIR1 insertions might significantly affect T-cell
463 epitopes, considering that the majority of T-cell response appears to be directed towards the spike NTD and
464 CTD domains (Tarke et al., 2021), cannot be ruled out. Comparative genomics investigations carried out on
465 other viruses belonging to the *Sarbecovirus* subgenus revealed that the RIR1 has been previously prone to
466 structural alterations during the radiation of bat coronaviruses (Garry et al., 2021). In fact, the spike
467 proteins RmYN02, RacCS203, BANAL-116 and BANAL-247 (Temmam et al., 2021; Zhou et al., 2021, 2020),
468 which are among the closest known relatives to SARS-CoV-2 when genomic recombination is taken into
469 account (Lytras et al., 2021), comprise an insertion of four codons at RIR1 in comparison with other bat
470 coronaviruses.
471 Most certainly, the spread of the A.2.5 and B.1.214.2 lineages in different geographical contexts between
472 late 2020 and early 2021, as well as the recent rapid global emergence of omicron, suggest that RIR1
473 insertions are unlikely to have a detrimental impact on the three-dimensional structure of the spike protein
474 or to significantly reduce the infectivity of these variants. At the same time, the well-defined length of the
475 insertions (in the overwhelming majority of cases 3 or 4 codons) suggests that some critical structural
476 constraints, that may prevent the selection of shorter insertions or limit their associated evolutionary
477 benefits, might exist. Several spike mutations located in the NTD can affect the structural organization of
478 the spike protein, altering the stability of the interaction between the RBD and the ACE2 receptor, or its
479 accessibility to antibody recognition. For instance, the NTD Δ69/Δ70 deletion, which, like RIR1, is found in
480 multiple independent lineages, does not determine a significant antibody escape *in vitro* (P. Wang et al.,
481 2021). However, it is thought to have an important impact on the structure of the spike protein, by
482 conferring increased cleavage at the S1/S2 site, thereby allowing higher incorporation in the virion (Meng
483 et al., 2021). In light of these observations, some NTD indels apparently not related with immune escape
484 may act as permissive mutations, by compensating small infectivity deficits associated with other RBD
485 mutations (i.e. L452R in A.2.5, Q414K and N450K in B.1.214.2, K417N, N440K, G446S, S477N, T478K, E484A,
486 Q493R, Q498R, N501Y and Y505H in omicron).
487 Interestingly, the insertion of a seven amino-acid long peptide at RIR1 in SARS-CoV-2 through passages in
488 Vero cell cultures has been recently implicated in enhanced *in vitro* infectivity, which may be linked with an
489 increase in the positive charge of NTD surface. According to the authors, this insertion (which is not
490 reported in Table 1 due to its laboratory origin) might have increased the affinity of the spike NTD to
491 heparin, bringing viral particles in close proximity with host cells, thereby favoring the interaction with
492 ACE2 (Shiliaev et al., 2021). While RIR1 insertions rarely share significant pairwise similarity both at the
493 nucleotide and at the amino acid level (**Figure 2, Table 1**), we tested whether the amino acids found in the

494 41 RIR1 insertions were over-represented compared to expectations (assuming that no codon usage bias
495 was present). As shown in **Supplementary Figure 1**, the basic amino acids arginine and lysine were the
496 most abundant ones (each one accounting for over 10% of total observations), followed by alanine,
497 glutamic acid and glycine. Overall, lysine was the amino acid characterized by the highest
498 observed/expected ratio (i.e. > 3) and also arginine showed a moderate increase in frequency compared to
499 expectations, supporting the conclusions by Shiliaev and colleagues about the benefits of acquiring basic
500 residues at RIR1 for viral infectivity. Nevertheless, the negatively charged glutamic acid was the second
501 most highly enriched amino acid and aspartic acid was also characterized by a positive observed/expected
502 ratio, which may extend such benefits to all charged residues. On the other hand, several amino acids with
503 hydrophobic side chains (e.g. I, M, T, V and Y in particular) were strongly under-represented. In general,
504 however, none only one out of the three widespread lineages with insertions at RIR1, i.e. B.1.214.2,
505 included a positively charged amino acid in the inserted tripeptide, suggesting that the hypothesis
506 proposed by Siliaev and colleagues might not be applicable to all insertions occurring at RIR1.
507 To preliminarily investigate the impact of RIR1 insertions on the structure of the spike protein, we applied
508 molecular dynamics simulations, a well-known technique able to capture and study the dynamical
509 properties of proteins and to assess the effects of mutations, deletions and insertions (Hansson et al.,
510 2002). In this case, we have used a coarse-grained force-field to compare the structural and/or dynamical
511 differences between the spike proteins from the wild-type virus and from the A.2.5 lineage. After 2µs
512 microseconds of simulations, the RMSD of the backbone atoms relative to the equivalent initial structures
513 (which represents a global measure of protein fluctuations) was calculated as a function of time to evaluate
514 the stability of MD simulations equilibrium in the two systems. No significant global displacement was
515 detected for any of the three protein models compared with the initial structure, as most of the RMSD
516 values only displayed fluctuations in a range between 0.35Å and 0.45Å. Similarly, the presence of a few
517 spike mutations in A.2.5 only led to minor changes in the compactness of the protein, as suggested by the
518 differences of about 1 Å found in the average RYGR values among the two models (**Supplementary Figure**
519 **2**). On the other hand, some fluctuations were visible in the RMSF of the A.2.5 spike protein model, in
520 particular in the regions which harbored non-synonymous mutations compared with the wild-type protein.
521 To understand changes in the direction of motions of the two systems under analysis, PCA was performed
522 on the last 2µs of the simulations, the time after which the systems reached the equilibration state. The
523 analysis was then restricted to the backbone beads, as they are less perturbed by statistical noise, providing
524 at the same time a significant characterization of the essential space motions (Michaud-Agrawal et al.,
525 2011). The diagonalization of the covariance matrix of fluctuations of the residues belonging to the
526 backbone resulted in a set of eigenvalues, which were plotted in decreasing order against the
527 corresponding eigenvector indices. The first few eigenvectors corresponded to concerted motions that
528 quickly decreased in amplitude to reach some constrained and more localized fluctuations. Here we
529 present the principal modes along the first eigenvector (**Figure 5A**), which covers about 25% of the motions
530 of the protein. Consistently with the placement of non-synonymous mutations (**Figure 3B**), this analysis
531 revealed that A.2.5 exhibited some changes in the fluctuations in regions belonging to the NTD and RBD,
532 which are shown in red in **Figure 5B**. This indicates that the presence of the mutations and insertions may
533 induce local structural and dynamical changes on the spike protein, highlighting the usefulness of
534 performing studies on the dynamical properties of insertions upon their emergence in variants with
535 widespread circulation.
536

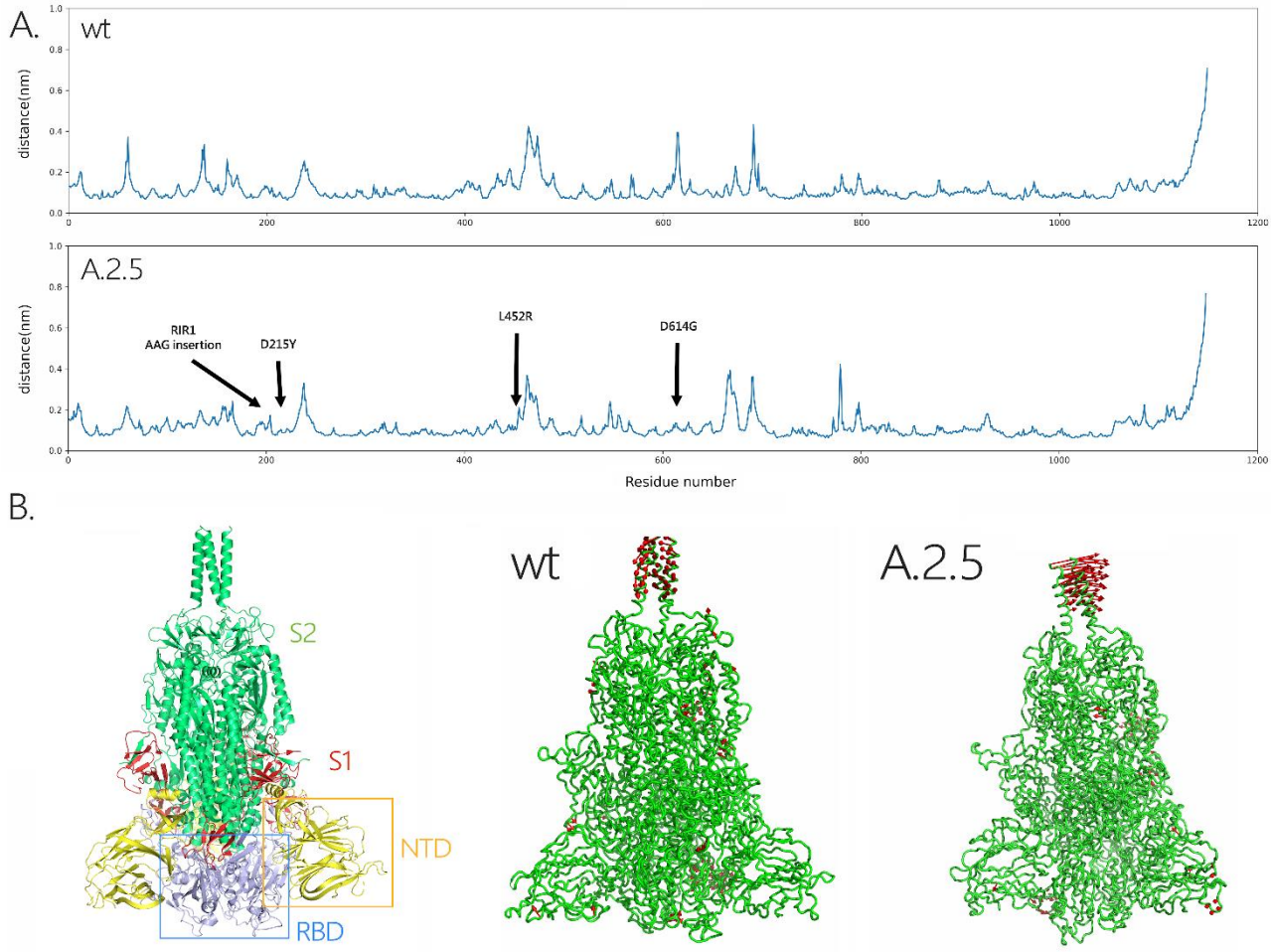


Figure 5: Panel A: RMSF plot for the models of the wild-type and A.2.5 SARS-CoV-2 spike proteins, with indication of the point and insertion mutations present in the two viral lineages target of his study, compared with the wild type virus. Panel B: Three-dimensional structural models obtained for the wild type and A.2.5 spike proteins. 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 are marked in red.

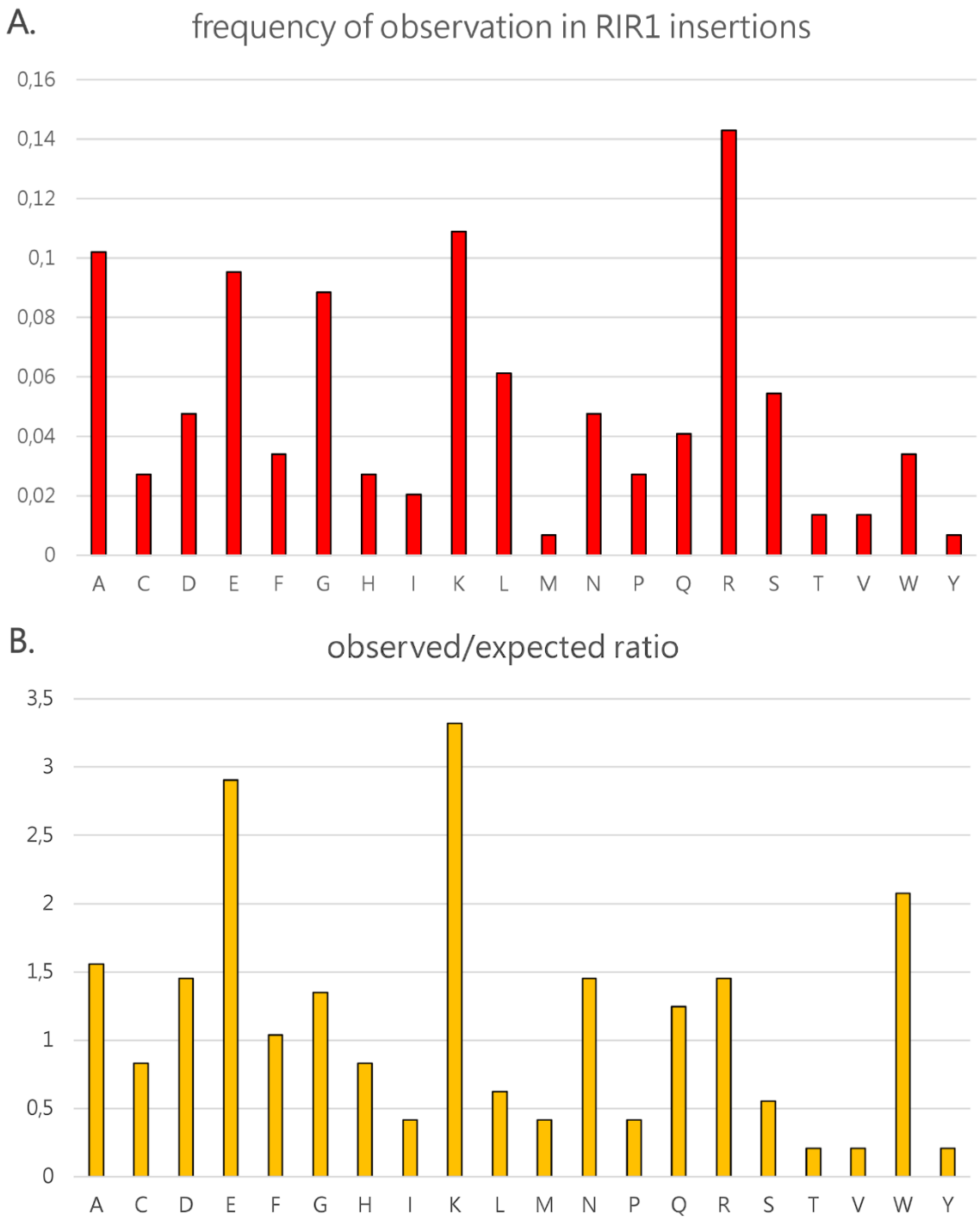
4. Conclusions

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. Prior to November 2021, the insertions at RIR1 documented in this work had only led to the emergence of two viral lineages with widespread transient distribution, i.e. B.1.214.2 and A.2.5. However, the presence of a RIR1 insertion in the emerging omicron VOC, together with the recurrent independent occurrence of this phenomenon by convergent evolution in multiple viral lineages, suggests that RIR1 insertions may be linked with an evolutionary advantage, whose magnitude is presently unclear.

In absence of functional data, the role of RIR1 insertions can be only speculated. Based on the lack of overlap with known immune epitopes, their involvement in immune escape phenomena appears unlikely, even though their impact on T-cell response remains to be investigated. Similarly, the previously hypothesized role of NTD insertions in enhancing viral infectivity by promoting the interaction with host cell membranes is only partly supported by the over-representation of lysine and arginine residues in RIR1 inserts. On the other hand, we observe a correlation between the presence of RIR1 insertions, RDR deletions and several non-synonymous RBD mutations with known impact on immune evasion on

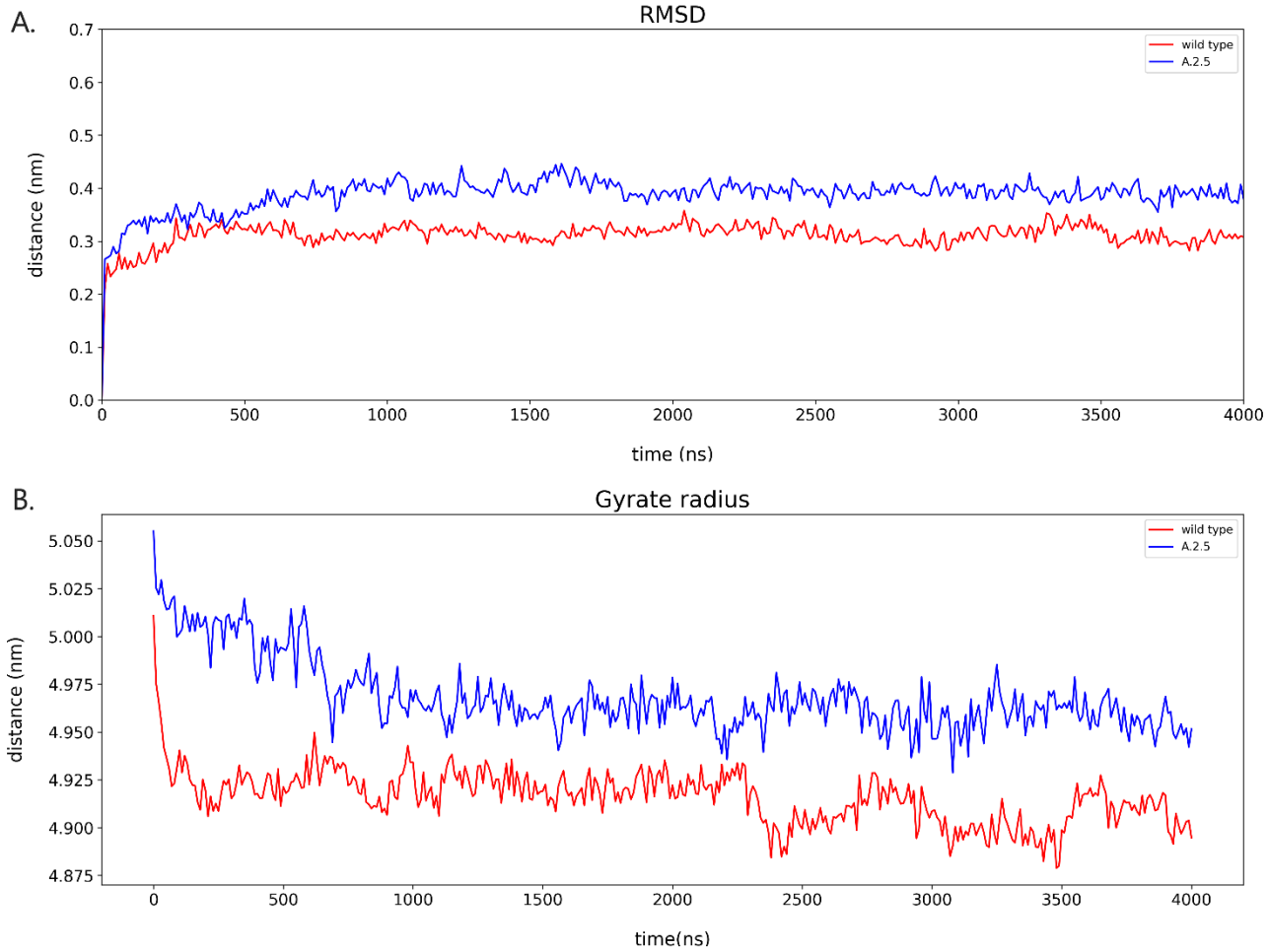
infectivity. This, together with the predicted effect of RIR1 insertions on the structure of the spike protein, may suggest a possible role as a permissive mutation in compensating otherwise slightly disadvantageous non-synonymous spike mutations. Undoubtedly, our observations strongly suggest that the functional and structural impact of these insertions should be the subject of in-depth studies in the near future.

Supplementary Figures



Supplementary Figure 1. Panel A: observed frequency of each amino acid in the 41 RIR1 insertions. Panel B: observed/expected ratios for each amino acid, calculated based under the assumption that no codon usage

570 bias was present. Amino acids showing a ratio > 1 were over-represented compared with expectations,
571 whereas those showing a ratio < 1 were under-represented.
572



573
574 **Supplementary Figure 2.** Panel A: RMSD plot of the three models of the SARS-CoV-2 spike protein (wild-type
575 and A.2.5), as a function of simulated time. The systems reach equilibrium after 1.5 µs. Panel B: variation of
576 RGYR observed over time for the three models of the SARS-CoV-2 spike protein (wild-type and A.2.5) during
577 the MD simulation.
578
579
580 **Acknowledgements**
581 The author is grateful to Dr. Alberto Beretta, Dr. Sarah Ann Nadeau and Dr. Alexander Martinez for their
582 communication and useful suggestions, and the staff of Pop Medicine
583 (https://www.facebook.com/medicipop/) for the support. The author also acknowledges all the
584 fundamental work carried out by the clinicians, researchers and public health authorities that allowed the
585 collection of SARS-CoV-2 genome data and made sequence data available in a timely manner though
586 GISAID (Lopez Bernal et al., 2021), as well as the great efforts made by the developers of nextstrain to assist
587 researchers in SARS-CoV-2 evolution studies.
588
589 **Competing Interest Statement**
590 The authors declare they have no competing interests.

Funding Statement

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Author contributions

MG: Conceptualization; Data curation; Formal analysis; Investigation; Methodology; Supervision; Visualization; Writing - original draft; Writing - review & editing. KD, AG: Formal analysis; Investigation; Writing - original draft; Writing - review & editing.

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

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. Emergence and spread of a SARS-CoV-2 lineage A variant (A.23.1) with altered spike protein in Uganda. Nat. Microbiol. 6, 1094–1101. https://doi.org/10.1038/s41564-021-00933-9

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

Díaz, Y., Ortiz, A., Weeden, A., Castillo, D., González, C., Moreno, B., Martínez-Montero, M., Castillo, M., Vasquez, G., Sáenz, L., Franco, D., Pitti, Y., Chavarria, O., Gondola, J., Moreno, A.M., Ábrego, L., Beltrán, D., Guerra, I., Chang, J., Chaverra, Z., Guerrero, I., Valoy, A., Gaitán, M., Araúz, D., Morán, E., Chen-Germán, M., Valdespino, E., Rodríguez, R., Corrales, R., Chen-Camaño, R., Pascale, J.M., Martínez, A.A., López-Vergès, S., 2021. SARS-CoV-2 reinfection with a virus harboring mutation in the Spike and the Nucleocapsid proteins in Panama. Int. J. Infect. Dis. 108, 588–591. https://doi.org/10.1016/j.ijid.2021.06.004

Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797. https://doi.org/10.1093/nar/gkh340

Ferreira, I.A.T.M., Kemp, S.A., Datir, R., Saito, A., Meng, B., Rakshit, P., Takaori-Kondo, A., Kosugi, Y., Uriu, K., Kimura, I., Shirakawa, K., Abdullahi, A., Agarwal, A., Ozono, S., Tokunaga, K., Sato, K., Gupta, R.K., CITIID-NIHR BioResource COVID-19 Collaboration, Indian SARS-CoV-2 Genomics Consortium, Genotype to Phenotype Japan (G2P-Japan) Consortium, 2021. SARS-CoV-2 B.1.617 Mutations L452R and E484Q Are Not Synergistic for Antibody Evasion. J. Infect. Dis. 224, 989–994. https://doi.org/10.1093/infdis/jiab368

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).

Garushyants, S.K., Rogozin, I.B., Koonin, E.V., 2021. Template switching and duplications in SARS-CoV-2 genomes give rise to insertion variants that merit monitoring. Commun. Biol. 4, 1–9. https://doi.org/10.1038/s42003-021-02858-9

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. Mapping mutations to the SARS-CoV-2 RBD that escape binding by different classes of antibodies. Nat. Commun. 12, 4196. https://doi.org/10.1038/s41467-021-24435-8

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

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

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

Kumar, S., Stecher, G., Li, M., Knyaz, C., Tamura, K., 2018. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. Mol. Biol. Evol. 35, 1547–1549. https://doi.org/10.1093/molbev/msy096

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. Infect. Genet. Evol. 95, 105038. https://doi.org/10.1016/j.meegid.2021.105038

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. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. Cell Host Microbe 29, 477-488.e4. https://doi.org/10.1016/j.chom.2021.01.014

Lopez Bernal, J., Andrews, N., Gower, C., Gallagher, E., Simmons, R., Thelwall, S., Stowe, J., Tessier, E., Groves, N., Dabrera, G., Myers, R., Campbell, C.N.J., Amirthalingam, G., Edmunds, M., Zambon, M., Brown, K.E., Hopkins, S., Chand, M., Ramsay, M., 2021. Effectiveness of Covid-19 Vaccines against the B.1.617.2 (Delta) Variant. N. Engl. J. Med. 385, 585–594. https://doi.org/10.1056/NEJMoa2108891

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., De Marco, A., Lempp, F.A., Tortorici, M.A., Pinto, D., Walls, A.C., Beltramello, M., Chen, A., Liu, Z., Zatta, F., Zepeda, S., di Iulio, J., 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. Cell 184, 2332-2347.e16. https://doi.org/10.1016/j.cell.2021.03.028

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

Meng, B., Kemp, S.A., Papa, G., Datir, R., Ferreira, I.A.T.M., Marelli, S., Harvey, W.T., Lytras, S., Mohamed, A., Gallo, G., Thakur, N., Collier, D.A., Mlcochova, P., Robson, S.C., Loman, N.J., Connor, T.R., Golubchik, T., Martinez Nunez, R.T., Ludden, C., Corden, S., Johnston, I., Bonsall, D., Smith, C.P., Awan, A.R., Bucca, G., Torok, M.E., Saeed, K., Prieto, J.A., Jackson, D.K., Hamilton, W.L., Snell, L.B., Moore, Catherine, Harrison, E.M., Goncalves, S., Fairley, D.J., Loose, M.W., Watkins, J., Livett, R., Moses, S., Amato, R., Nicholls, S., Bull, M., Smith, D.L., Barrett, Jeff, Aanensen, D.M., Curran, M.D., Parmar, S., Aggarwal, D., Shepherd, J.G., Parker, M.D., Glaysher, S., Bashton, M., Underwood, A.P., Pacchiarini, N., Loveson, K.F., Templeton, K.E., Langford, C.F., Sillitoe, J., de Silva, T.I., Wang, D., Kwiatkowski, D., Rambaut, A., O'Grady, J., Cottrell, S., Holden, M.T.G., Thomson, E.C., Osman, H., Andersson, M., Chauhan, A.J., Hassan-Ibrahim, M.O., Lawniczak, M., Alderton, A., Chand, M., Constantinidou, C., Unnikrishnan, M., Darby, A.C., Hiscox, J.A., Paterson, S., Martincorena, I., Volz, E.M., Page, A.J., Pybus, O.G., Bassett, A.R., Ariani, C.V., Chapman, M.H.S., Li, K.K., Shah, R.N., Jesudason, N.G., Taha, Y., McHugh, M.P., Dewar, R., Jahun, A.S., McMurray, C., Pandey, S., McKenna, J.P., Nelson, A., Young, G.R., McCann, C.M., Elliott, S., Lowe, H., Temperton, B., Roy, S., Price, A., Rey, S., Wyles, M., Rooke, S., Shaaban, S., de Cesare, M., Letchford, L., Silveira, S., Pelosi, E., Wilson-Davies, E., Hosmillo, M., O'Toole, Á., Hesketh, A.R., Stark, R., du Plessis, L., Ruis, C., Adams, H., Bourgeois, Y., Michell, S.L., Gramatopoulos, D., Edgeworth, J., Breuer, J., Todd, J.A., Fraser, C., Buck, D., John, M., Kay, G.L., Palmer, Steve, Peacock, S.J., Heyburn, D., Weldon, D., Robinson, E., McNally, A., Muir, P., Vipond, I.B., Boyes, J., Sivaprakasam, V., Salluja, T., Dervisevic, S., Meader, E.J., Park, N.R., Oliver, K., Jeffries, A.R., Ott, S., da Silva Filipe, A., Simpson, D.A., Williams, C., Masoli, J.A.H., Knight, B.A., Jones, C.R., Koshy, C., Ash, A., Casey, A., Bosworth, A., Ratcliffe, L., Xu-McCrae, L., Pymont, H.M., Hutchings, S., Berry, Lisa, Jones, K., Halstead, F., Davis, T., Holmes, C., Iturriza-Gomara, M., Lucaci, A.O., Randell, P.A., Cox, A., Madona, P., Harris, K.A., Brown, J.R., Mahungu, T.W., Irish-Tavares, D., Haque, T., Hart, J., Witele, E., Fenton, M.L., Liggett, S., Graham, C., Swindells, E., Collins, J., Eltringham, G., Campbell, S., McClure, P.C., Clark, G., Sloan, T.J., Jones, C., Lynch, J., Warne, B., Leonard, S., Durham, J., Williams, T., Haldenby, S.T., Storey, N., Alikhan, N.-F., Holmes, N., Moore, Christopher, Carlile, M., Perry, M., Craine, N., Lyons, R.A., Beckett, A.H., Goudarzi, S., Fearn, C., Cook, K., Dent, H., Paul, H., Davies, R., Blane, B., Girgis, S.T., Beale, M.A., Bellis, K.L., Dorman, M.J., Drury, E., Kane, L., Kay, S., McGuigan, S., Nelson, R., Prestwood, L., Rajatileka, S., Batra, R., Williams, R.J., Kristiansen, M., Green, A., Justice, A., Mahanama, A.I.K., Samaraweera, B., Hadjirin, N.F., Quick, J., Poplawski, R., Kermack, L.M., Reynolds, N., Hall, G., Chaudhry, Y., Pinckert, M.L., Georgana, I., Moll, R.J., Thornton, A., Myers, R.,

Stockton, J., Williams, C.A., Yew, W.C., Trotter, A.J., Trebes, A., MacIntyre-Cockett, G., Birchley, A., Adams, A., Plimmer, A., Gatica-Wilcox, B., McKerr, C., Hilvers, E., Jones, H., Asad, H., Coombes, J., Evans, J.M., Fina, L., Gilbert, L., Graham, L., Cronin, M., Kumziene-Summerhayes, S., Taylor, Sarah, Jones, S., Groves, D.C., Zhang, P., Gallis, M., Louka, S.F., Starinskij, I., Jackson, C., Gourtovaia, M., Tonkin-Hill, G., Lewis, K., Tovar-Corona, J.M., James, K., Baxter, L., Alam, M.T., Orton, R.J., Hughes, J., Vattipally, S., Ragonnet-Cronin, M., Nascimento, F.F., Jorgensen, D., Boyd, O., Geidelberg, L., Zarebski, A.E., Raghwani, J., Kraemer, M.U.G., Southgate, J., Lindsey, B.B., Freeman, T.M., Keatley, J.-P., Singer, J.B., de Oliveira Martins, L., Yeats, C.A., Abudahab, K., Taylor, B.E.W., Menegazzo, M., Danesh, J., Hogsden, W., Eldirdiri, S., Kenyon, A., Mason, J., Robinson, T.I., Holmes, A., Price, J., Hartley, J.A., Curran, T., Mather, A.E., Shankar, G., Jones, R., Howe, R., Morgan, S., Wastenge, E., Chapman, M.R., Mookerjee, S., Stanley, Rachael, Smith, W., Peto, T., Eyre, D., Crook, D., Vernet, G., Kitchen, C., Gulliver, H., Merrick, I., Guest, M., Munn, R., Bradley, D.T., Wyatt, T., Beaver, C., Foulser, L., Palmer, Sophie, Churcher, C.M., Brooks, E., Smith, K.S., Galai, K., McManus, G.M., Bolt, F., Coll, F., Meadows, L., Attwood, S.W., Davies, A., De Lacy, E., Downing, F., Edwards, S., Scarlett, G.P., Jeremiah, S., Smith, N., Leek, D., Sridhar, S., Forrest, S., Cormie, C., Gill, H.K., Dias, J., Higginson, E.E., Maes, M., Young, J., Wantoch, M., Jamrozy, D., Lo, S., Patel, M., Hill, V., Bewshea, C.M., Ellard, S., Auckland, C., Harrison, I., Bishop, C., Chalker, V., Richter, A., Beggs, A., Best, A., Percival, B., Mirza, J., Megram, O., Mayhew, M., Crawford, L., Ashcroft, F., Moles-Garcia, E., Cumley, N., Hopes, R., Asamaphan, P., Niebel, M.O., Gunson, R.N., Bradley, A., Maclean, A., Mollett, G., Blacow, R., Bird, P., Helmer, T., Fallon, K., Tang, J., Hale, A.D., Macfarlane-Smith, L.R., Harper, K.L., Carden, H., Machin, N.W., Jackson, K.A., Ahmad, S.S.Y., George, R.P., Turtle, L., O'Toole, E., Watts, J., Breen, C., Cowell, A., Alcolea-Medina, A., Charalampous, T., Patel, A., Levett, L.J., Heaney, J., Rowan, A., Taylor, G.P., Shah, D., Atkinson, L., Lee, J.C.D., Westhorpe, A.P., Jannoo, R., Lowe, H.L., Karamani, A., Ensell, L., Chatterton, W., Pusok, M., Dadrah, A., Symmonds, A., Sluga, G., Molnar, Z., Baker, P., Bonner, S., Essex, S., Barton, E., Padgett, D., Scott, G., Greenaway, J., Payne, B.A.I., Burton-Fanning, S., Waugh, S., Raviprakash, V., Sheriff, N., Blakey, V., Williams, L.-A., Moore, J., Stonehouse, S., Smith, L., Davidson, R.K., Bedford, L., Coupland, L., Wright, V., Chappell, J.G., Tsoleridis, T., Ball, J., Khakh, M., Fleming, V.M., Lister, M.M., Howson-Wells, H.C., Berry, Louise, Boswell, T., Joseph, A., Willingham, I., Duckworth, N., Walsh, S., Wise, E., Moore, N., Mori, M., Cortes, N., Kidd, S., Williams, R., Gifford, L., Bicknell, K., Wyllie, S., Lloyd, A., Impey, R., Malone, C.S., Cogger, B.J., Levene, N., Monaghan, L., Keeley, A.J., Partridge, D.G., Raza, M., Evans, C., Johnson, K., Betteridge, E., Farr, B.W., Goodwin, S., Quail, M.A., Scott, C., Shirley, L., Thurston, S.A.J., Rajan, D., Bronner, I.F., Aigrain, L., Redshaw, N.M., Lensing, S.V., McCarthy, S., Makunin, Alex, Balcazar, C.E., Gallagher, M.D., Williamson, K.A., Stanton, T.D., Michelsen, M.L., Warwick-Dugdale, J., Manley, R., Farbos, A., Harrison, J.W., Sambles, C.M., Studholme, D.J., Lackenby, A., Mbisa, T., Platt, S., Miah, S., Bibby, D., Manso, C., Hubb, J., Dabrera, G., Ramsay, M., Bradshaw, D., Schaefer, U., Groves, N., Gallagher, E., Lee, D., Williams, D., Ellaby, N., Hartman, H., Manesis, N., Patel, V., Ledesma, J., Twohig, K.A., Allara, E., Pearson, C., Cheng, J.K.J., Bridgewater, H.E., Frost, L.R., Taylor-Joyce, G., Brown, P.E., Tong, L., Broos, A., Mair, D., Nichols, J., Carmichael, S.N., Smollett, K.L., Nomikou, K., Aranday-Cortes, E., Johnson, N., Nickbakhsh, S., Vamos, E.E., Hughes, M., Rainbow, L., Eccles, R., Nelson, C., Whitehead, M., Gregory, R., Gemmell, M., Wierzbicki, C., Webster, H.J., Fisher, C.L., Signell, A.W., Betancor, G., Wilson, H.D., Nebbia, G., Flaviani, F., Cerda, A.C., Merrill, T.V., Wilson, R.E., Cotic, M., Bayzid, N., Thompson, T., Acheson, E., Rushton, S., O'Brien, S., Baker, D.J., Rudder, S., Aydin, A., Sang, F., Debebe, J., Francois, S., Vasylyeva, T.I., Zamudio, M.E., Gutierrez, B., Marchbank, A., Maksimovic, J., Spellman, K., McCluggage, K., Morgan, M., Beer, R., Afifi, S., Workman, T., Fuller, W., Bresner, C., Angyal, A., Green, L.R., Parsons, P.J., Tucker, R.M., Brown, R., Whiteley, M., Bonfield, J., Puethe, C., Whitwham, A., Liddle, Jennifier, Rowe, W., Siveroni, I., Le-Viet, T., Gaskin, A., Johnson, R., Abnizova, I., Ali, M., Allen, L., Anderson, R., Ariani, C., Austin-Guest, S., Bala, S., Barrett, Jeffrey, Bassett, A., Battleday, K., Beal, J., Beale, M., Bellany, S., Bellerby, T., Bellis, K., Berger, D., Berriman, M., Bevan, P., Binley, S., Bishop, J., Blackburn, K., Boughton, N., Bowker, S., Brendler-Spaeth, T., Bronner, I., Brooklyn, T., Buddenborg, S.K., Bush, R., Caetano, C., Cagan, A., Carter, N., Cartwright, J., Monteiro, T.C., Chapman, L., Chillingworth, T.-J., Clapham, P.,

Clark, R., Clarke, A., Clarke, C., Cole, D., Cook, E., Coppola, M., Cornell, L., Cornwell, C., Corton, C., Crackett, A., Cranage, A., Craven, H., Craw, S., Crawford, M., Cutts, T., Dabrowska, M., Davies, M., Dawson, J., Day, C., Densem, A., Dibling, T., Dockree, C., Dodd, D., Dogga, S., Dorman, M., Dougan, G., Dougherty, M., Dove, A., Drummond, L., Dudek, M., Durrant, L., Easthope, E., Eckert, S., Ellis, P., Farr, B., Fenton, M., Ferrero, M., Flack, N., Fordham, H., Forsythe, G., Francis, M., Fraser, A., Freeman, A., Galvin, A., Garcia-Casado, M., Gedny, A., Girgis, S., Glover, J., Gould, O., Gray, A., Gray, E., Griffiths, C., Gu, Y., Guerin, F., Hamilton, W., Hanks, H., Harrison, E., Harrott, A., Harry, E., Harvison, J., Heath, P., Hernandez-Koutoucheva, A., Hobbs, R., Holland, D., Holmes, S., Hornett, G., Hough, N., Huckle, L., Hughes-Hallet, L., Hunter, A., Inglis, S., Iqbal, S., Jackson, A., Jackson, D., Verdejo, C.J., Jones, M., Kallepally, K., Kay, K., Keatley, J., Keith, A., King, A., Kitchin, L., Kleanthous, M., Klimekova, M., Korlevic, P., Krasheninnkova, K., Lane, G., Langford, C., Laverack, A., Law, K., Lensing, S., Lewis-Wade, A., Liddle, Jennifer, Lin, Q., Lindsay, S., Linsdell, S., Long, R., Lovell, Jamie, Lovell, Jon, Mack, J., Maddison, M., Makunin, Aleksei, Mamun, I., Mansfield, J., Marriott, N., Martin, M., Mayho, M., McClintock, J., McHugh, S., MapcMinn, L., Meadows, C., Mobley, E., Moll, R., Morra, M., Morrow, L., Murie, K., Nash, S., Nathwani, C., Naydenova, P., Neaverson, A., Nerou, E., Nicholson, J., Nimz, T., Noell, G.G., O'Meara, S., Ohan, V., Olney, C., Ormond, D., Oszlanczi, A., Pang, Y.F., Pardubska, B., Park, N., Parmar, A., Patel, G., Payne, M., Peacock, S., Petersen, A., Plowman, D., Preston, T., Quail, M., Rance, R., Rawlings, S., Redshaw, N., Reynolds, J., Reynolds, M., Rice, S., Richardson, M., Roberts, C., Robinson, K., Robinson, M., Robinson, D., Rogers, H., Rojo, E.M., Roopra, D., Rose, M., Rudd, L., Sadri, R., Salmon, N., Saul, D., Schwach, F., Seekings, P., Simms, A., Sinnott, M., Sivadasan, S., Siwek, B., Sizer, D., Skeldon, K., Skelton, J., Slater-Tunstill, J., Sloper, L., Smerdon, N., Smith, Chris, Smith, Christen, Smith, J., Smith, K., Smith, M., Smith, S., Smith, T., Sneade, L., Soria, C.D., Sousa, C., Souster, E., Sparkes, A., Spencer-Chapman, M., Squares, J., Stanley, Robert, Steed, C., Stickland, T., Still, I., Stratton, M., Strickland, M., Swann, A., Swiatkowska, A., Sycamore, N., Swift, E., Symons, E., Szluha, S., Taluy, E., Tao, N., Taylor, K., Taylor, Sam, Thompson, S., Thompson, M., Thomson, M., Thomson, N., Thurston, S., Toombs, D., Topping, B., Tovar-Corona, J., Ungureanu, D., Uphill, J., Urbanova, J., Van, P.J., Vancollie, V., Voak, P., Walker, D., Walker, M., Waller, M., Ward, G., Weatherhogg, C., Webb, N., Wells, A., Wells, E., Westwood, L., Whipp, T., Whiteley, T., Whitton, G., Widaa, S., Williams, M., Wilson, M., Wright, S., Duncan, L.M., Carabelli, A.M., Kenyon, J.C., Lever, A.M., De Marco, A., Saliba, C., Culap, K., Cameroni, E., Matheson, N.J., Piccoli, L., Corti, D., James, L.C., Robertson, D.L., Bailey, D., Gupta, R.K., 2021. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. Cell Rep. 35, 109292. https://doi.org/10.1016/j.celrep.2021.109292

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-Osorno, M., Brenes, H., Soto-Garita, C., Pérez-Corrales, C., COINGESA-CR Consorcio Interinstitucional de Estudios Genómicos del SARS-CoV-2 Costa Rica, Drexler, J.F., Moreira-Soto, A., Corrales-Aguilar, E., Duarte-Martínez, F., 2021. SARS-CoV-2 genomic surveillance in Costa Rica: Evidence of a divergent population and an increased detection of a spike T1117I mutation. Infect. Genet. Evol. J. Mol. Epidemiol. Evol. Genet. Infect. Dis. 92, 104872. https://doi.org/10.1016/j.meegid.2021.104872

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

Ozer, E.A., Simons, L.M., Adewumi, O.M., Fowotade, A.A., Omoruyi, E.C., Adeniji, J.A., Dean, T.J., Zayas, J., Bhimalli, P.P., Ash, M.K., Godzik, A., Schneider, J.R., Mamede, J.I., Taiwo, B.O., Hultquist, J.F., Lorenzo-Redondo, R., 2021. Coincident rapid expansion of two SARS-CoV-2 lineages with enhanced

infectivity in Nigeria. medRxiv 2021.04.09.21255206. https://doi.org/10.1101/2021.04.09.21255206

Peacock, T.P., Bauer, D.L., Barcley, W.S., 2021. Putative host origins of RNA insertions in SARS-CoV-2 genomes - SARS-CoV-2 coronavirus. Virological.

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., Le Guen, J., Kassis-Chikhani, N., Edriss, D., Belec, L., Seve, A., Courtellemont, L., Péré, H., Hocqueloux, L., Fafi-Kremer, S., Prazuck, T., Mouquet, H., Bruel, T., Simon-Lorière, E., Rey, F.A., Schwartz, O., 2021. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature 596, 276–280. https://doi.org/10.1038/s41586-021-03777-9

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

Pulliam, J.R.C., Schalkwyk, C. van, Govender, N., Gottberg, A. von, Cohen, C., Groome, M.J., Dushoff, J., Mlisana, K., Moultrie, H., 2021. Increased risk of SARS-CoV-2 reinfection associated with emergence of the Omicron variant in South Africa. https://doi.org/10.1101/2021.11.11.21266068

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., da Rocha, A.S.B., Nascimento, V., Souza, V., Silva, G., Nascimento, F., Neto, L.G.L., da Silva, F.V., Riediger, I., Debur, M. do C., Leite, A.B., Mattos, T., da Costa, C.F., Pereira, F.M., dos Santos, C.A., Rovaris, D.B., Fernandes, S.B., Abbud, A., Sacchi, C., Khouri, R., Bernardes, A.F.L., Delatorre, E., Gräf, T., Siqueira, M.M., Bello, G., Wallau, G.L., on behalf of Fiocruz COVID-19 Genomic Surveillance Network, 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. Virus Evol. 7. https://doi.org/10.1093/ve/veab069

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

Shiliaev, N., Lukash, T., Palchevska, O., Crossman, D.K., Green, T.J., Crowley, M.R., Frolova, E.I., Frolov, I., 2021. Natural and Recombinant SARS-CoV-2 Isolates Rapidly Evolve In Vitro to Higher Infectivity through More Efficient Binding to Heparan Sulfate and Reduced S1/S2 Cleavage. J. Virol. 95, e01357-21. https://doi.org/10.1128/JVI.01357-21

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., 2021. 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

Tarke, A., Sidney, J., Kidd, C.K., Dan, J.M., Ramirez, S.I., Yu, E.D., Mateus, J., da Silva Antunes, R., Moore, E., Rubiro, P., Methot, N., Phillips, E., Mallal, S., Frazier, A., Rawlings, S.A., Greenbaum, J.A., Peters, B., Smith, D.M., Crotty, S., Weiskopf, D., Grifoni, A., Sette, A., 2021. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. Cell Rep. Med. 2, 100204. https://doi.org/10.1016/j.xcrm.2021.100204

Temmam, S., Vongphayloth, K., Salazar, E.B., Munier, S., Bonomi, M., Régnault, B., Douangboubpha, B., Karami, Y., Chretien, D., Sanamxay, D., Xayaphet, V., Paphaphanh, P., Lacoste, V., Somlor, S., Lakeomany, K., Phommavanh, N., Pérot, P., Donati, F., Bigot, T., Nilges, M., Rey, F., Werf, S. van der, Brey, P., Eloit, M., 2021. Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese peninsula. https://doi.org/10.21203/rs.3.rs-871965/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., 2021. Modelling conformational state dynamics and its role on infection for SARS-CoV-2 Spike protein variants. PLOS Comput. Biol. 17, e1009286. https://doi.org/10.1371/journal.pcbi.1009286

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

Venkatakrishnan, A.J., Anand, P., Lenehan, P.J., Suratekar, R., Raghunathan, B., Niesen, M.J.M., Soundararajan, V., 2021. Omicron variant of SARS-CoV-2 harbors a unique insertion mutation of putative viral or human genomic origin. https://doi.org/10.31219/osf.io/f7txy

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.,

1000 Stokes, M.M., Hemachudha, T., Wang, L.-F., 2021. Evidence for SARS-CoV-2 related coronaviruses
1001 circulating in bats and pangolins in Southeast Asia. Nat. Commun. 12, 972.
1002 https://doi.org/10.1038/s41467-021-21240-1
1003 Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, Function, and
1004 Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 181, 281-292.e6.
1005 https://doi.org/10.1016/j.cell.2020.02.058
1006 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,
1007 M.-J., 2021. Susceptibility of Circulating SARS-CoV-2 Variants to Neutralization. N. Engl. J. Med. 0,
1008 null. https://doi.org/10.1056/NEJMc2103022
1009 Wang, P., Nair, M.S., Liu, L., Iketani, S., Luo, Y., Guo, Y., Wang, M., Yu, J., Zhang, B., Kwong, P.D., Graham,
1010 B.S., Mascola, J.R., Chang, J.Y., Yin, M.T., Sobieszczyk, M., Kyratsous, C.A., Shapiro, L., Sheng, Z.,
1011 Huang, Y., Ho, D.D., 2021. Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7. Nature
1012 1–6. https://doi.org/10.1038/s41586-021-03398-2
1013 Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finkin, S., Schaefer-Babajew, D., Cipolla, M.,
1014 Gaebler, C., Lieberman, J.A., Oliveira, T.Y., Yang, Z., Abernathy, M.E., Huey-Tubman, K.E., Hurley, A.,
1015 Turroja, M., West, K.A., Gordon, K., Millard, K.G., Ramos, V., Da Silva, J., Xu, J., Colbert, R.A., Patel,
1016 R., Dizon, J., Unson-O'Brien, C., Shimeliovich, I., Gazumyan, A., Caskey, M., Bjorkman, P.J., Casellas,
1017 R., Hatziioannou, T., Bieniasz, P.D., Nussenzweig, M.C., 2021. mRNA vaccine-elicited antibodies to
1018 SARS-CoV-2 and circulating variants. Nature 592, 616–622. https://doi.org/10.1038/s41586-021-
1019 03324-6
1020 Waterhouse, A., Bertoni, M., Bienert, S., Studer, G., Tauriello, G., Gumienny, R., Heer, F.T., de Beer, T.A.P.,
1021 Rempfer, C., Bordoli, L., Lepore, R., Schwede, T., 2018. SWISS-MODEL: homology modelling of
1022 protein structures and complexes. Nucleic Acids Res. 46, W296–W303.
1023 https://doi.org/10.1093/nar/gky427
1024 Xie, X., Liu, Y., Liu, J., Zhang, X., Zou, J., Fontes-Garfias, C.R., Xia, H., Swanson, K.A., Cutler, M., Cooper, D.,
1025 Menachery, V.D., Weaver, S.C., Dormitzer, P.R., Shi, P.-Y., 2021. Neutralization of SARS-CoV-2 spike
1026 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat. Med. 27, 620–
1027 621. https://doi.org/10.1038/s41591-021-01270-4
1028 Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Peng, H., Quinlan, B.D., Rangarajan, E.S., Pan, A., Vanderheiden,
1029 A., Suthar, M.S., Li, W., Izard, T., Rader, C., Farzan, M., Choe, H., 2020. SARS-CoV-2 spike-protein
1030 D614G mutation increases virion spike density and infectivity. Nat. Commun. 11, 6013.
1031 https://doi.org/10.1038/s41467-020-19808-4
1032 Zhang, W., Davis, B.D., Chen, S.S., Sincuir Martinez, J.M., Plummer, J.T., Vail, E., 2021. Emergence of a Novel
1033 SARS-CoV-2 Variant in Southern California. JAMA 325, 1324–1326.
1034 https://doi.org/10.1001/jama.2021.1612
1035 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.,
1036 Shi, W., 2020. A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at
1037 the S1/S2 Cleavage Site of the Spike Protein. Curr. Biol. 30, 2196-2203.e3.
1038 https://doi.org/10.1016/j.cub.2020.05.023
1039 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,
1040 A.C., Holmes, E.C., Shi, W., 2021. Identification of novel bat coronaviruses sheds light on the
1041 evolutionary origins of SARS-CoV-2 and related viruses. Cell.
1042 https://doi.org/10.1016/j.cell.2021.06.008
1043 Zhu, X., Mannar, D., Srivastava, S.S., Berezuk, A.M., Demers, J.-P., Saville, J.W., Leopold, K., Li, W., Dimitrov,
1044 D.S., Tuttle, K.S., Zhou, S., Chittori, S., Subramaniam, S., 2021. Cryo-electron microscopy structures
1045 of the N501Y SARS-CoV-2 spike protein in complex with ACE2 and 2 potent neutralizing antibodies.
1046 PLOS Biol. 19, e3001237. https://doi.org/10.1371/journal.pbio.3001237
1047