CORONAVIRUS

Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape

Kevin R. McCarthy[1,2,3]*, Linda J. Rennick[1,2], Sham Nambulli[1,2], Lindsey R. Robinson-McCarthy[4], William G. Bain[5,6,7], Ghady Haidar[8,9], W. Paul Duprex[1,2]*

Zoonotic pandemics, such as that caused by severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), can follow the spillover of animal viruses into highly susceptible human populations. The descendants of these viruses have adapted to the human host and evolved to evade immune pressure. Coronaviruses acquire substitutions more slowly than other RNA viruses. In the spike glycoprotein, we found that recurrent deletions overcome this slow substitution rate. Deletion variants arise in diverse genetic and geographic backgrounds, transmit efficiently, and are present in novel lineages, including those of current global concern. They frequently occupy recurrent deletion regions (RDRs), which map to defined antibody epitopes. Deletions in RDRs confer resistance to neutralizing antibodies. By altering stretches of amino acids, deletions appear to accelerate SARS-CoV-2 antigenic evolution and may, more generally, drive adaptive evolution.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) emerged from a yet-to-be-defined animal reservoir and initiated a pandemic in 2020 (*1*–*5*). It has acquired limited adaptions, most notably the Asp^{614} → Gly (D614G) substitution in the spike (S) glycoprotein (*6*–*8*). Humoral immunity to S glycoprotein appears to be the strongest correlate of protection (*9*), and recently approved vaccines deliver this antigen by immunization. Coronaviruses such as SARS-CoV-2 acquire substitutions slowly as the result of a proofreading RNA-dependent RNA polymerase (RdRp) (*10*, *11*). Other emerging respiratory viruses have produced pandemics followed by endemic human-to-human spread. The latter is often contingent upon the introduction of antigenic novelty that enables reinfection of previously immune individuals. Whether SARS-CoV-2 S glycoprotein will evolve altered antigenicity, or specifically how it may change in response to immune pressure, remains unknown. We and others have reported the acquisition of deletions in the N-terminal domain (NTD) of the S glycoprotein during long-term infections of immunocompromised patients (*12*–*15*). We have identified this as an evolutionary pattern defined by recurrent deletions that alter defined antibody epitopes. Unlike substitutions, deletions cannot be corrected by proofreading activity, and this may accelerate adaptive evolution in SARS-CoV-2.

An immunocompromised cancer patient infected with SARS-CoV-2 was unable to clear the virus and succumbed to the infection 74 days after COVID-19 diagnosis (*15*). Treatment included remdesivir, dexamethasone, and two infusions of convalescent serum. We designate this individual as Pittsburgh long-term infection 1 (PLTI1). We consensus-sequenced and cloned S genes directly from clinical material obtained 72 days after COVID-19 diagnosis and identified two variants with deletions in the NTD (Fig. 1A).

These data from PLTI1 and a similar report (*12*) prompted us to interrogate patient metadata sequences deposited in GISAID (*16*). In searching for similar viruses, we identified eight patients with deletions in the S glycoproteins of viruses sampled longitudinally over a period of weeks to months (Fig. 1A and fig. S1A). For each, early time points had intact S sequences and later time points had deletions within the S gene. Six had deletions that were identical to, overlapping with, or adjacent to those in PLTI1. Deletions at a second site were present in viruses isolated from two other patients (Fig. 1B); reports on these patients have since been published (*13*, *14*). Viruses from all but one patient could be distinguished from one another by nucleotide differences present at both early and late time points (fig. S1B). On a tree of representative contemporaneously circulating isolates, they form monophyletic clades, making either a second community-acquired or nosocomially acquired infection unlikely (fig. S1C). The most parsimonious explanation is that these deletions arose independently as the result of a common selective pressure to produce strikingly convergent outcomes.

We searched the GISAID sequence database (*16*) for additional instances of deletions within S glycoproteins. From a dataset of 146,795 sequences (deposited from 1 December 2019 to 24 October 2020) we identified 1108 viruses with deletions in the S gene. When mapped to the S gene, 90% of these deletions occupied four discrete sites within the NTD (Fig. 2A). We term these important sites recurrent deletion regions (RDRs), numbering them 1 to 4 from the 5′ to the 3′ end of the S gene. Deletions identified in patient samples correspond to RDR2 (Fig. 1A) and RDR4 (Fig. 1B). Most deletions appear to have arisen and been retained in replication-competent viruses.

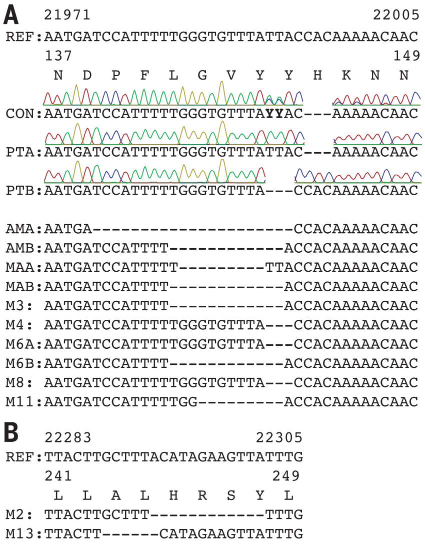

A
21971
22005
REF:AATGATCCATTTTTGGGTGTTTATTACCACAAAAACAAC
137
149
N D P F L G V Y Y H K N N
CON:AATGATCCATTTTTGGGTGTTTAYYAC---AAAAACAAC
PTA:AATGATCCATTTTTGGGTGTTTATTAC---AAAAACAAC
PTB:AATGATCCATTTTTGGGTGTTTA---CCACAAAAACAAC
AMA:AATGA---------------------CCACAAAAACAAC
AMB:AATGATCCATTTT------------ACCACAAAAACAAC
MAA:AATGATCCATTTTT---------TTACCACAAAAACAAC
MAB:AATGATCCATTTT------------ACCACAAAAACAAC
M3: AATGATCCATTTT------------ACCACAAAAACAAC
M4: AATGATCCATTTTTGGGTGTTTA---CCACAAAAACAAC
M6A:AATGATCCATTTTTGGGTGTTTA---CCACAAAAACAAC
M6B:AATGATCCATTTT------------ACCACAAAAACAAC
M8: AATGATCCATTTTTGGGTGTTTA---CCACAAAAACAAC
M11:AATGATCCATTTTTGG---------ACCACAAAAACAAC
B
22283
22305
REF:TTACTTGCTTTACATAGAAGTTATTTG
241
249
L L A L H R S Y L
M2: TTACTTGCTTT------------TTTG
M13:TTACTT------CATAGAAGTTATTTG

Fig. 1. Deletions in SARS-CoV-2 spike glycoprotein arise during persistent infections of immunosuppressed patients. (**A**) Top: Sequences of viruses isolated from PLTI1 (PT) and viruses from patients with deletions in the same NTD region. Chromatograms are shown for sequences from PLTI1, which include sequencing of bulk reverse transcription products (CON) and individual cDNA clones. Bottom: Sequences from other long-term infections from individuals AM (*18*), MA-JL (MA) (*19*), and a MSK cohort (M) with individuals 3, 4, 6, 8, and 11 (*13*). Letters (A and B) designate different variants from the same patient. (**B**) Sequences of viruses from two patients (M2 and M13) with deletions in a different region of the NTD. All sequences are aligned to reference sequence (REF) MN985325 (WA-1). See fig. S1 for genetic analysis of patient isolates. Amino acid abbreviations: A, Ala; D, Asp; F, Phe; G, Gly; H, His; K, Lys; L, Leu; N, Asn; P, Pro; R, Arg; S, Ser; V, Val; Y, Tyr.

[1] Center for Vaccine Research, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [2] Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [3] Laboratory of Molecular Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. [4] Department of Genetics, Harvard Medical School, Boston, MA, USA. [5] Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Internal Medicine, UPMC, Pittsburgh, PA, USA. [6] Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [7] VA Pittsburgh Healthcare System, Pittsburgh, PA, USA. [8] Division of Infectious Disease, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. [9] Division of Infectious Disease, Department of Internal Medicine, UPMC, Pittsburgh, PA, USA.
***Corresponding author. Email: krm@pitt.edu (K.R.M.); pduprex@pitt.edu (W.P.D.)**

Without selective pressure, in-frame deletions should occur one-third of the time. However, we observed a preponderance of in-frame deletions with lengths of 3, 6, 9, and 12 (Fig. 2B). Among all deletions, 93% are in frame and do not produce a stop codon (Fig. 2C). In the NTD, >97% of deletions maintain the open reading frame. Other S glycoprotein domains do not follow this trend; for example, deletions in the receptor binding domain (RBD) and S2 preserve the reading frame 30% and 37% of the time, respectively.

To trace the origins of RDR variants, we produced phylogenies for each with 101 additional genomes that sample much of the genetic diversity within the pandemic (Fig. 2D). The RDR variants interleave with nondeletion sequences and occupy distinct branches, indicating their recurrent generation. This is most pronounced for RDRs 1, 2, and 4 but is also true of RDR3, with conservatively four independent instances. RDR variants form distinct lineages/branches, most prominently in RDR1 (lineage B.1.258), and suggest human-to-human transmission events. Using sequences with sufficient metadata to explicitly differentiate individuals, we verified the transmission of a variant within each RDR between people (fig. S2).

We defined the RDRs on the basis of peaks in the spectrum of S glycoprotein deletions. Deletion lengths and positions vary within RDRs 1, 2, and 4 (Fig. 2E). Variation is greatest in RDRs 2 and 4, with the loss of S glycoprotein residues 144/145 (adjacent tyrosine codons) in RDR2 and residues 243 and 244 in RDR4 appearing to be favored. In contrast, the loss of residues 69 and 70 accounts for the vast majority of RDR1 deletions. On the basis of our phylogenetic analysis and accompanying lineage classifications, this two–amino acid deletion has arisen independently at least 13 times. RDR3 largely consists of three nucleotide deletions in codon 220.

We evaluated the genetic, geographic, and temporal sampling of RDR variants (Fig. 3, A and B). This analysis was limited to sequences deposited in GISAID (*16*) where sequences from specific nations and regions are overrepresented (e.g., United Kingdom and other European countries). We show the distribution of all sequences within the database for reference. For RDR2 and RDR4, the genetic and geographic distributions largely mirror those of reported sequences. Variants of RDR1 and RDR3 are strongly polarized to specific clades and geographies. This is likely the result of successful lineages circulating in regions with strong sequencing initiatives. Our temporal analysis indicates that RDR variants have been present throughout the pandemic (Fig. 3C). Specific variant lineages such as B.1.258 (Fig. 2D) harboring Δ69–70 in RDR1 have rapidly risen to notable abundance (Fig. 3D). Circulation of B.1.36 with RDR3 Δ210 accounts for most of the RDR3 examples (Fig. 2D and Fig. 3, C and D). The abundance of RDR2 Δ144/145 is explained by independent deletion events followed by transmission (Fig. 2D and Fig. 3, C and D).

The recurrence and convergence of RDR deletions, particularly during long-term infections, is indicative of adaptation in response to a common selective pressure. RDRs 2 and 4 and RDRs 1 and 3 occupy two distinct surfaces on the S glycoprotein NTD (Fig. 4A). Both sites contain antibody epitopes (*17*–*19*). The epitope for neutralizing antibody 4A8 is formed entirely by the β sheets and extended connecting loops that harbor RDRs 2 and 4 (*17*). We generated a panel of S glycoprotein mutants representing the four RDRs to assess the impact of deletions on expression and antibody binding; we included an additional double mutant containing the deletions present in the B.1.1.7 variant of concern flagged initially in the United Kingdom. Cells were transfected with plasmids expressing these mutant glycoproteins, and indirect immunofluorescence was used to determine whether RDR deletions modulated 4A8 binding (Fig. 4B). Deletions at RDRs 1 and 3 had no impact on the binding of the monoclonal antibody, confirming that they alter independent sites. The three RDR2 deletions, the one RDR4 deletion, and the double RDR1/2 deletions completely abolished binding of 4A8 while still allowing recognition by a monoclonal antibody

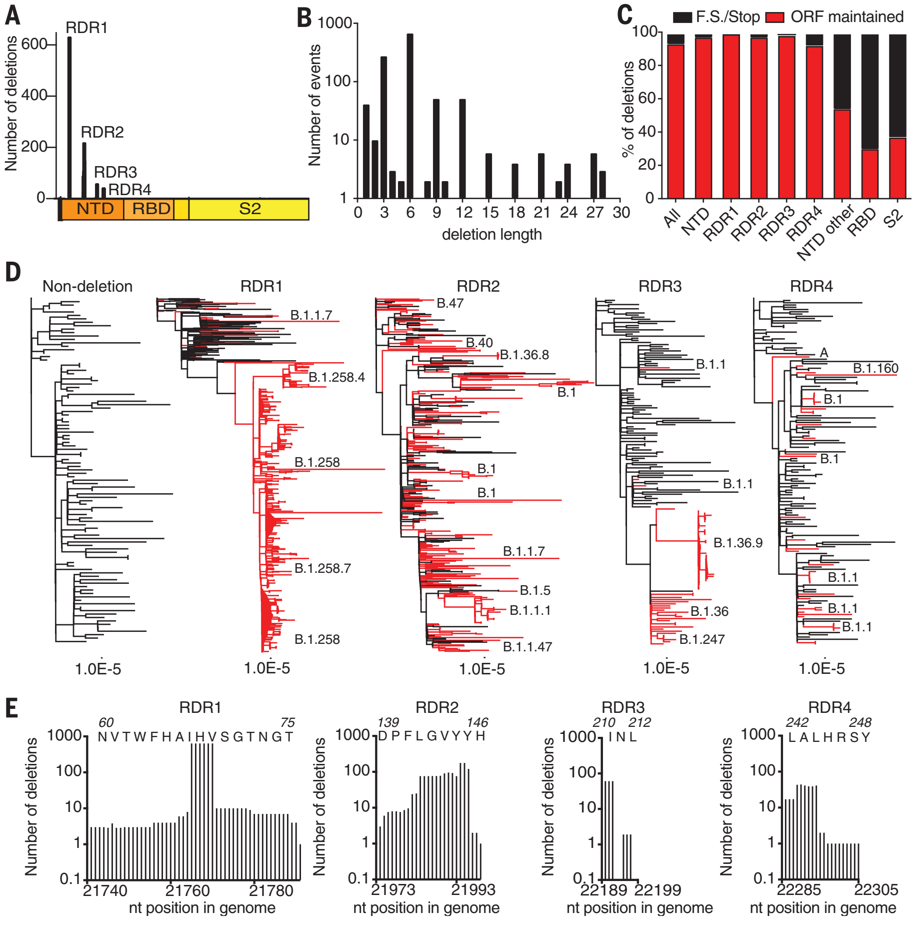


Fig. 2. Identification and characterization of recurrent deletion regions in SARS-CoV-2 spike protein. (**A**) Positional quantification of deleted nucleotides in S among GISAID sequences. We designate the four clusters as recurrent deletion regions (RDRs) 1 to 4. (**B**) Length distribution of deletions. (**C**) The percentage of deletion events at the indicated site that either maintain the open reading frame (ORF) or introduce a frameshift or premature stop codon (F.S./Stop). (**D**) Phylogenetic analysis of deletion variants (red branches) and genetically diverse nondeletion variants (black branches). Specific deletion clades/lineages are identified. Maximum likelihood phylogenetic trees, rooted on NC_045512, were calculated with 1000 bootstrap replicates. Trees with branch labels are in fig. S2. (**E**) Abundance of nucleotide (nt) deletions in each RDR. Positions are defined by reference sequence MN985325, by codon (top) and nucleotide (below). Amino acid abbreviations: A, Ala; D, Asp; F, Phe; G, Gly; H, His; I, Ile; L, Leu; N, Asn; P, Pro; R, Arg; S, Ser; T, Thr; V, Val; W, Trp; Y, Tyr.

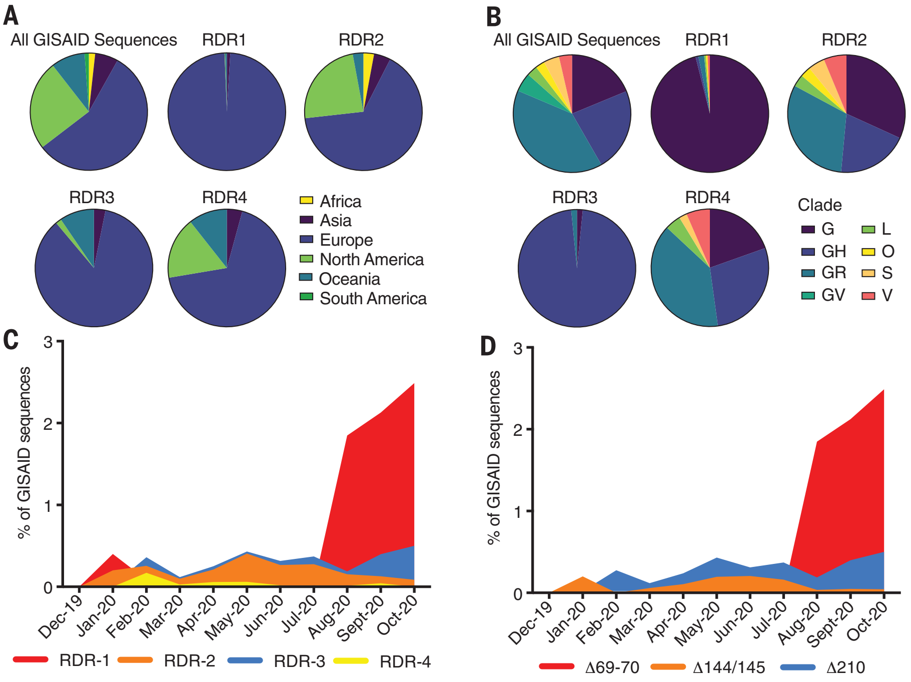
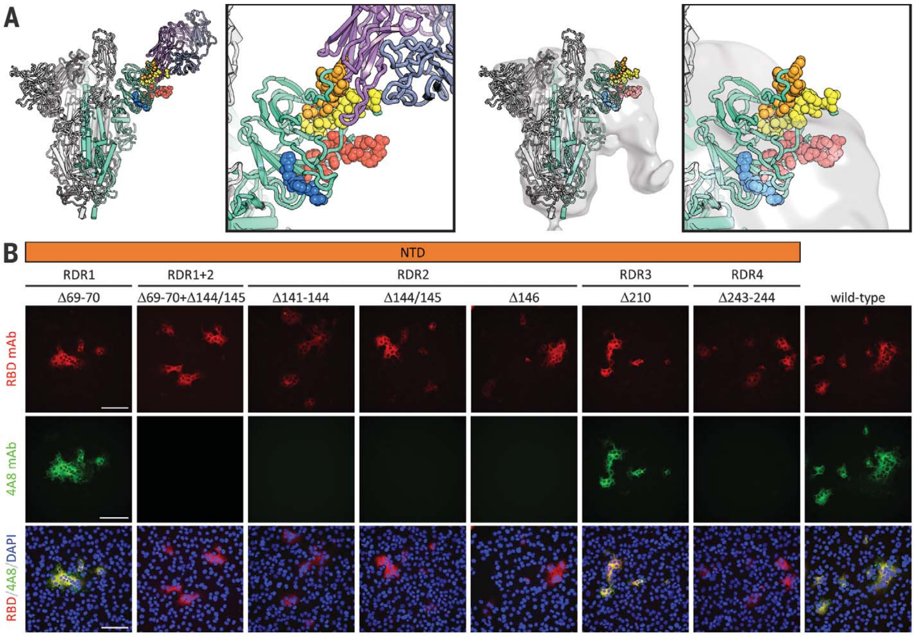


Fig. 3. Geographic, genetic, and temporal abundance of RDR variants. (**A** and **B**) Geographic (A) and genetic (B) distributions of RDR variants compared to the GISAID database (sequences from 1 December 2019 to 24 October 2020). GISAID clade classifications are used in (B). (**C**) Frequency of RDR variants among all complete genomes deposited in GISAID. (**D**) Frequency of specific RDR deletion variants (numbered according to spike amino acids) among all GISAID variants. The plot of RDR3/Δ210 has been adjusted by 0.02 units on the *y* axis for visualization in (C) because of its overlap with RDR2, and this adjustment has been retained in (D) to enable direct comparisons between panels.



C

antibody	virus	25	12.5	6.25	3.125	1.56	0.78	0.39	0.2	0.1	0.05
		antibody concentration (μg/ml)									
4A8 (α-S)	PLTI1	-	-	-	-	-	-	-	-	-	-
	Munich: P3	+	+	+	+	+	+	-	-	-	-
H2214 (α-HA)	PLTI1	-	-	-	-	-	-	ND	ND	ND	ND
	Munich: P3	-	-	-	-	-	-	ND	ND	ND	ND
	virus	serum dilution									
		1:100	1:200	1:400	1:800	1:1600	1:3200				
human convalescent serum	PLTI1	+	+	+	+	+	-				
	Munich: P3	+	+	+	+	+	-				

+ neutralization (by PRNT 80)
- no neutralization (by PRNT 80)
ND not determined

Fig. 4. Deletions in the spike NTD alter its antigenicity; RDRs map to defined antigenic sites. (**A**) Left: A structure of antibody 4A8 (*17*) (PDB ID 7C21) (purple) bound to one protomer (green) of a SARS-CoV-2 spike trimer (gray). RDRs 1 to 4 are colored red, orange, blue, and yellow, respectively, and are shown as spheres. The boxed image is a close-up of the interaction site. Right: The electron microscopy density of COV57 serum Fabs (*18*) (EMDB emd_22125) fit to SARS-CoV-2 S glycoprotein trimer (PDB ID 7C21). The boxed image is a close-up of the interaction site. (**B**) S glycoprotein distribution in Vero E6 cells at 24 hours after transfection with S protein deletion mutants, visualized by indirect immunofluorescence in permeabilized cells. A monoclonal antibody to SARS-CoV-2 S protein receptor-binding domain (RBD mAb; red) detects all mutant forms of the protein (Δ69–70, Δ69–70+Δ141–144, Δ141–144, Δ144/145, Δ146, Δ210, and Δ243–244) and the unmodified protein (wild type), whereas 4A8 mAb (green) does not detect mutants containing deletions in RDR2 or RDR4 (Δ69–70+Δ141–144, Δ141–144, Δ144/145, Δ146, and Δ243–244). Overlay images (RBD/4A8/DAPI) depict colocalization of the antibodies; nuclei were counterstained with 4′,6-diamidino-2-phenylindole (DAPI; blue). Scale bars, 100 μm. (**C**) Virus isolated from PLTI1 resists neutralization by 4A8. A nondeletion variant (Munich) is neutralized by 4A8, both are neutralized by convalescent serum, and neither is neutralized by H2214, an influenza hemagglutinin binding antibody (*29*).

targeting the RBD (Fig. 4B). Thus, convergent evolution operates in individual RDRs and between RDRs, as exemplified by the same phenotype produced by deletions in RDR2 or RDR4.

We used the non–plaque-purified viral population from PLTI1 to determine whether RDR variants escape the activity of a neutralizing antibody. This viral stock was completely resistant to neutralization by 4A8, whereas an isolate with authentic RDRs (*20*) was neutralized (Fig. 4C). We used a high-titer neutralizing human convalescent polyclonal antiserum to demonstrate that both viral stocks could be neutralized efficiently. These data demonstrate that naturally arising and circulating variants of SARS-CoV-2 have altered antigenicity. We used a range of high-, medium-, and low-titer neutralizing human convalescent polyclonal antisera to assess whether there was an appreciable difference in neutralization between the S glycoprotein–deleted and undeleted viruses. No major difference was observed, which suggests that many more changes would be required to generate serologically distinct SARS-CoV-2 variants (table S1).

Coronaviruses, including SARS-CoV-2, have lower substitution rates than other RNA viruses because of an RdRp with proofreading activity (*10*, *11*). However, proofreading cannot correct deletions. We find that adaptive evolution of S glycoprotein is augmented by a tolerance for deletions, particularly within RDRs. The RDRs occupy defined antibody epitopes within the NTD (*17*–*19*), and deletions at multiple sites confer resistance to a neutralizing antibody. Deletions represent a generalizable mechanism through which S glycoprotein rapidly acquires genetic and antigenic novelty of SARS-CoV-2.

The fitness of RDR variants is evident by their representation in the consensus genomes from patients, transmission between individuals, and presence in emergent lineages. Initially documented in the context of long-term infections of immunosuppressed patients, specific variants transmit efficiently between immunocompetent individuals. Characterization of these cases led to the very early identification of RDR variants that are escape mutants. Because deletions are a product of replication, they will occur at a certain rate and variants are likely to emerge in otherwise healthy populations. Indeed, influenza explores variation that approximates future antigenic drift in immunosuppressed patients (*21*).

The RDRs occupy defined antibody epitopes within the S glycoprotein NTD. Selected in vivo, these deletion variants resist neutralization by monoclonal antibodies. Viruses cultured in vitro in the presence of immune serum have also acquired substitutions in RDR2 that confer neutralization resistance (*22*). Potent neutralizing responses and an array of monoclonal antibodies are directed to the RBD (*18*, *19*, *23*). A growing number of NTD-directed antibodies have been identified (*24*, *25*). Why antibody escape in nature is most evident in the NTD highlights a discrepancy, and this requires further study.

Defining recurrent, convergent patterns of adaptation can provide predictive potential. From viral sequences, we have identified a pattern of deletions, contextualized their outcomes in protein structure and antibody epitope(s), and characterized their functional impact on antigenicity. During evaluation of this manuscript, multiple lineages with altered antigenicity and perhaps increased transmissibility have emerged and spread. These variants of global concern are RDR variants and include Mink Cluster 5 Δ69–70 (*26*), B.1.1.7 Δ69–70, and Δ144/145 (*27*), as well as B.1.351 Δ242–244 (*28*). Our analysis preceded the description of these lineages. We had demonstrated that identical or similar recurrent deletions that alter positions 144/145 and 243–244 in the S glycoprotein disrupt binding of antibody 4A8, which defines an immunodominant epitope within the NTD. Our survey for deletion variants captured the first representative of what would become the B.1.1.7 lineage. These real-world outcomes demonstrate the predictive potential of this and like approaches and show the need to monitor viral evolution carefully and continually.

Additional circulating RDR variants have gone virtually unnoticed. Are they intermediates on a pathway of immune evasion? That remains to be determined. However, deletions and substitutions within major NTD and RBD epitopes will likely continue to contribute to that process, as they have already in current variants of concern. The progression of adaptations in both immunocompromised patients and SARS-CoV-2 variants of concern remains to be resolved. Their evolution has thus far converged. The recurrence of adaptations in single patients and on global scales underscores the need to track and monitor deletion variants.

REFERENCES AND NOTES

1. N. Zhu *et al.*, *N. Engl. J. Med.* **382**, 727–733 (2020).
2. F. Wu *et al.*, *Nature* **579**, 265–269 (2020).
3. H. Zhou *et al.*, *Curr. Biol.* **30**, 3896 (2020).
4. T. T. Lam *et al.*, *Nature* **583**, 282–285 (2020).
5. M. F. Boni *et al.*, *Nat. Microbiol.* **5**, 1408–1417 (2020).
6. B. Korber *et al.*, *Cell* **182**, 812–827.e19 (2020).
7. R. P. McNamara *et al.*, *Cell Rep.* **33**, 108352 (2020).
8. E. Volz *et al.*, *Cell* **184**, 64–75.e11 (2021).
9. K. McMahan *et al.*, *Nature* 10.1038/s41586-020-03041-6 (2020).
10. M. R. Denison, R. L. Graham, E. F. Donaldson, L. D. Eckerle, R. S. Baric, *RNA Biol.* **8**, 270–279 (2011).
11. E. Minskaia *et al.*, *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5108–5113 (2006).
12. V. A. Avanzato *et al.*, *Cell* **183**, 1901–1912.e9 (2020).
13. T. Aydillo *et al.*, *N. Engl. J. Med.* **383**, 2586–2588 (2020).
14. B. Choi *et al.*, *N. Engl. J. Med.* **383**, 2291–2293 (2020).
15. M. K. Hensley *et al.*, *Clin. Infect. Dis.* ciab072 (2021).
16. Y. Shu, J. McCauley, *Euro Surveill.* **22**, 30494 (2017).
17. X. Chi *et al.*, *Science* **369**, 650–655 (2020).
18. C. O. Barnes *et al.*, *Cell* **182**, 828–842.e16 (2020).
19. L. Liu *et al.*, *Nature* **584**, 450–456 (2020).
20. W. B. Klimstra *et al.*, *J. Gen. Virol.* **101**, 1156–1169 (2020).
21. K. S. Xue *et al.*, *eLife* **6**, e26875 (2017).
22. Y. Weisblum *et al.*, *eLife* **9**, e61312 (2020).
23. L. Piccoli *et al.*, *Cell* **183**, 1024–1042.e21 (2020).
24. D. Li *et al.*, bioRxiv 10.1101/2020.12.31.424729 [preprint]. 2 January 2021; https://www.biorxiv.org/content/10.1101/2020.12.31.424729v1.
25. W. N. Voss *et al.*, bioRxiv 10.1101/2020.12.20.423708 [preprint]. 21 December 2020; https://www.biorxiv.org/content/10.1101/2020.12.20.423708v1.
26. World Health Organization, "SARS-CoV-2 mink-associated variant strain – Denmark" (2020); www.who.int/csr/don/06-november-2020-mink-associated-sars-cov2-denmark/en/.
27. S. A. Kemp *et al.*, bioRxiv 10.1101/2020.12.14.422555 [preprint]. 21 December 2020; https://www.biorxiv.org/content/10.1101/2020.12.14.422555v2.
28. H. Tegally *et al.*, medRxiv 10.1101/2020.12.21.20248640 [preprint]. 22 December 2020; https://www.medrxiv.org/content/10.1101/2020.12.21.20248640v1.
29. A. Watanabe *et al.*, *Cell* **177**, 1124–1135.e16 (2019).

ACKNOWLEDGMENTS

We gratefully acknowledge the authors from the originating laboratories and the submitting laboratories, who generated and shared via GISAID genetic sequence data on which this research is based (table S2). We thank S. C. Harrison for his support. We thank A. Morris, B. McVerry, G. Kitsios, B. Methe, J. S. Lee, P. Ray, H. Michael, M. Busch, J. Ries, and C. Schaefer at the University of Pittsburgh, as well as the physicians, nurses, and respiratory therapists at the University of Pittsburgh Medical Center Shadyside-Presbyterian Hospital intensive care units for assistance with collection and processing of the endotracheal aspirate sample. This work was conducted under University of Pittsburgh Review Board approval CR19050099-099. **Author contributions:** K.R.M., L.J.R., S.N., L.R.R.-M., and W.P.D. designed the experiments; K.R.M., L.J.R., S.N., and L.R.R.M. performed the experiments; K.R.M., L.J.R., S.N., L.R.R.-M., and W.P.D. analyzed data; W.G.B. and G.H. provided reagents and samples; and K.R.M., L.J.R., S.N., L.R.R.-M., and W.P.D. wrote the manuscript. **Competing interests:** The authors declare no competing interests. **Funding:** Supported by the University of Pittsburgh, the Center for Vaccine Research, the Commonwealth of Pennsylvania Department of Community and Economic Development, the Richard King Mellon Foundation, the Henry L. Hillman Foundation (W.P.D.), National Heart, Lung, and Blood Institute award P01HL114453 and U.S. Department of Veterans Affairs Career Development Award IK2 BX004886 (W.G.B.), and UPMC Immune Transplant and Therapy Center (W.G.B., G.H.). **Data availability:** Sequences from PLTI1 were deposited in NCBI GenBank under accession numbers MW269404 and MW269555. All other sequences are available via the GISAID SARS-CoV-2 sequence database (www.gsaid.org; see table S2 for full list of acknowledgments).

SUPPLEMENTARY MATERIALS

science.sciencemag.org/content/371/6534/1139/suppl/DC1
Materials and Methods
Figs. S1 to S3
Tables S1 and S2
References (*30*–*34*)
MDAR Reproducibility Checklist

19 November 2020; accepted 30 January 2021
Published online 3 February 2021
10.1126/science.abf6950