

Cite as: K. R. McCarthy *et al.*, *Science*
10.1126/science.abf6950 (2021).

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-McCarthy⁴, William G. Bain^{5,6,7}, Ghady Haidar^{8,9}, W. Paul Duprex^{1,2*}

¹Center for Vaccine Research, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ²Department of Microbiology and Molecular Genetics, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ³Laboratory of Molecular Medicine, Boston Children's Hospital, Harvard Medical School, Boston, MA, USA. ⁴Department of Genetics, Harvard Medical School, Boston, MA, USA. ⁵Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Internal Medicine, UPMC, Pittsburgh, PA, USA. ⁶Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ⁷VA Pittsburgh Healthcare System, Pittsburgh, PA, USA. ⁸Division of Infectious Disease, Department of Medicine, University of Pittsburgh School of Medicine, Pittsburgh, PA, USA. ⁹Division 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.)

Zoonotic pandemics, like that caused by SARS-CoV-2, can follow the spillover of animal viruses into highly susceptible human populations. Their descendants have adapted to the human host and evolved to evade immune pressure. Coronaviruses acquire substitutions more slowly than other RNA viruses, due to a proofreading polymerase. In the spike glycoprotein, we find 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.

SARS-CoV-2 emerged from a yet-to-be defined animal reservoir and initiated a pandemic in 2020 (1–5). It has acquired limited adaptations, most notably the 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 like SARS-CoV-2 slowly acquire substitutions due to 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 amino (N)-terminal domain (NTD) of the S glycoprotein during long-term infections of often-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 the individual as Pittsburgh long-term infection 1 (PLTI1). We consensus sequenced and cloned S genes directly from clinical material obtained 72 days following 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- or nosocomially-acquired infection unlikely (fig. S1C). The most parsimonious explanation

is these deletions arose independently due to 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 12/01/2019 to 10/24/2020) we identified 1,108 viruses with deletions in the S gene. When mapped to the S gene, 90% occupied four discrete sites within the NTD (Fig. 2A). We term these important sites recurrent deletion regions (RDRs), numbering them 1-4 from the 5' to 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. 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 e.g., 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 non-deletion sequences and occupy distinct branches, indicating their recurrent generation. This is most pronounced for RDRs 1, 2 and 4 but also true of RDR 3, 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. We verified, using sequences with sufficient metadata that explicitly differentiate individuals, the transmission of a variant within each RDR between people (fig. S2).

We defined the RDRs based upon 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 243-244 in RDR4 appearing to be favored. In contrast, the loss of residues 69-70 accounts for the vast majority of RDR1 deletions. Based upon our phylogenetic analysis and supported by accompanying lineage classifications this two amino acid deletion has arisen independently at least thirteen times. RDR3 largely consists of three nucleotide (nt) deletions in codon 220.

We evaluated the genetic, geographic and temporal sampling of RDR variants (Fig. 3, A and B). This analysis is limited to sequences deposited in GISAID (16) where sequences from specific nations and regions are overrepre-

sented e.g., United Kingdom and other European countries. We show the distribution of all sequences within the database for reference. For RDRs 2 and 4 the genetic and geographic distributions largely mirror those of reported sequences. Variants of RDRs 1 and 3 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 like B.1.258 (Fig. 2D) harboring $\Delta 69-70$ in RDR1 have rapidly risen to notable abundance (Fig. 3D). Circulation of B.1.36 with RDR3 $\Delta 210$ accounts for most of the RDR3 examples (Figs. 2D and Fig. 3, C and D). The abundance of RDR2 $\Delta 144/145$ 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 beta 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 deletions have 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 if 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 targeting the RBD (Fig. 4B). Thus, convergent evolution operates in individual RDRs and between RDRs, exemplified by the same phenotype produced by deletions in RDR2 or RDR4.

We assayed whether RDR variants escape the activity of a neutralizing antibody using the non-plaque purified viral population from PLTI1. This viral stock was completely resistant to neutralization by 4A8, while 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 as-

sess if there was an appreciable difference in neutralization between the S glycoprotein-deleted and undeleted viruses. No major difference was observed suggesting 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 due to 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.

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 unique cases led to the very early identification of RDR variants that are escape mutants. Since 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 identify a pattern of deletions, contextualize their outcomes in protein structure and antibody epitope(s), and characterize 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 Δ 69-70 (26), B.1.1.7 Δ 69-70 and Δ 144/145 (27), and B.1.351 Δ 242-244 (28). Our analysis preceded the description of these lineages. We had demonstrated that identical/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 immunocompromised patients and concerning SARS-CoV-2 variants alike remain to be resolved. Their evolution has thus far converged. 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, D. Zhang, W. Wang, X. Li, B. Yang, J. Song, X. Zhao, B. Huang, W. Shi, R. Lu, P. Niu, F. Zhan, X. Ma, D. Wang, W. Xu, G. Wu, G. F. Gao, W. Tan, A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* 382, 727–733 (2020). [doi:10.1056/NEJMoa2001017](https://doi.org/10.1056/NEJMoa2001017) [Medline](#)
2. F. Wu, S. Zhao, B. Yu, Y.-M. Chen, W. Wang, Z.-G. Song, Y. Hu, Z.-W. Tao, J.-H. Tian, Y.-Y. Pei, M.-L. Yuan, Y.-L. Zhang, F.-H. Dai, Y. Liu, Q.-M. Wang, J.-J. Zheng, L. Xu, E. C. Holmes, Y.-Z. Zhang, A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269 (2020). [doi:10.1038/s41586-020-2008-3](https://doi.org/10.1038/s41586-020-2008-3) [Medline](#)
3. H. Zhou, X. Chen, T. Hu, J. Li, H. Song, Y. Liu, P. Wang, D. Liu, J. Yang, E. C. Holmes, A. C. Hughes, Y. Bi, W. Shi, A Novel Bat Coronavirus Closely Related to SARS-CoV-2 Contains Natural Insertions at the S1/S2 Cleavage Site of the Spike Protein. *Curr. Biol.* 30, 3896 (2020). [doi:10.1016/j.cub.2020.09.030](https://doi.org/10.1016/j.cub.2020.09.030) [Medline](#)
4. T. T. Lam, N. Jia, Y.-W. Zhang, M. H.-H. Shum, J.-F. Jiang, H.-C. Zhu, Y.-G. Tong, Y.-X. Shi, X.-B. Ni, Y.-S. Liao, W.-J. Li, B.-G. Jiang, W. Wei, T.-T. Yuan, K. Zheng, X.-M. Cui, J. Li, G.-Q. Pei, X. Qiang, W. Y.-M. Cheung, L.-F. Li, F.-F. Sun, S. Qin, J.-C. Huang, G. M. Leung, E. C. Holmes, Y.-L. Hu, Y. Guan, W.-C. Cao, Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. *Nature* 583, 282–285 (2020). [doi:10.1038/s41586-020-2169-0](https://doi.org/10.1038/s41586-020-2169-0) [Medline](#)
5. M. F. Boni, P. Lemey, X. Jiang, T. T.-Y. Lam, B. W. Perry, T. A. Castoe, A. Rambaut, D. L. Robertson, Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat. Microbiol.* 5, 1408–1417 (2020). [doi:10.1038/s41564-020-0771-4](https://doi.org/10.1038/s41564-020-0771-4) [Medline](#)
6. B. Korber, W. M. Fischer, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, K. M. Hastie, M. D. Parker, D. G. Partridge, C. M. Evans, T. M. Freeman, T. I. de Silva, C. McDanal, L. G. Perez, H. Tang, A. Moon-Walker, S. P. Whelan, C. C. LaBranche, E. O. Saphire, D. C. Montefiori, A. Angyal, R. L. Brown, L. Carrilero, L. R. Green, D. C. Groves, K. J. Johnson, A. J. Keeley, B. B. Lindsey, P. J. Parsons, M. Raza, S. Rowland-Jones, N. Smith, R. M. Tucker, D. Wang, M. D. Wyles, Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* 182, 812–827.e19 (2020). [doi:10.1016/j.cell.2020.06.043](https://doi.org/10.1016/j.cell.2020.06.043) [Medline](#)
7. R. P. McNamara, C. Caro-Vegas, J. T. Landis, R. Moorad, L. J. Pluta, A. B. Eason, C. Thompson, A. Bailey, F. C. S. Villamor, P. T. Lange, J. P. Wong, T. Seltzer, J. Seltzer, Y. Zhou, W. Vahrson, A. Juarez, J. O. Meyo, T. Calabre, G. Broussard, R. Rivera-Soto, D. L. Chappell, R. S. Baric, B. Damania, M. B. Miller, D. P. Dittmer,

- High-Density Amplicon Sequencing Identifies Community Spread and Ongoing Evolution of SARS-CoV-2 in the Southern United States. *Cell Rep.* **33**, 108352 (2020). [doi:10.1016/j.celrep.2020.108352](https://doi.org/10.1016/j.celrep.2020.108352) [Medline](#)
8. E. Volz, V. Hill, J. T. McCrone, A. Price, D. Jorgensen, Á. O'Toole, J. Southgate, R. Johnson, B. Jackson, F. F. Nascimento, S. M. Rey, S. M. Nicholls, R. M. Colquhoun, A. da Silva Filipe, J. Shepherd, D. J. Pascall, R. Shah, N. Jesudason, K. Li, R. Jarrett, N. Pacchiarini, M. Bull, L. Geidelberg, I. Siveroni, COG-UK Consortium, I. Goodfellow, N. J. Loman, O. G. Pybus, D. L. Robertson, E. C. Thomson, A. Rambaut, T. R. Connor, Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **184**, 64–75.e11 (2021). [doi:10.1016/j.cell.2020.11.020](https://doi.org/10.1016/j.cell.2020.11.020) [Medline](#)
 9. K. McMahan, J. Yu, N. B. Mercado, C. Loos, L. H. Tostanoski, A. Chandrashekar, J. Liu, L. Peter, C. Atyeo, A. Zhu, E. A. Bondzie, G. Dagotto, M. S. Gebre, C. Jacob-Dolan, Z. Li, F. Nampanya, S. Patel, L. Pessaint, A. Van Ry, K. Blade, J. Yalley-Ogunro, M. Cabus, R. Brown, A. Cook, E. Teow, H. Andersen, M. G. Lewis, D. A. Lauffenburger, G. Alter, D. H. Barouch, Correlates of protection against SARS-CoV-2 in rhesus macaques. *Nature* **10.1038/s41586-020-03041-6** (2020). [doi:10.1038/s41586-020-03041-6](https://doi.org/10.1038/s41586-020-03041-6) [Medline](#)
 10. M. R. Denison, R. L. Graham, E. F. Donaldson, L. D. Eckerle, R. S. Baric, Coronaviruses: An RNA proofreading machine regulates replication fidelity and diversity. *RNA Biol.* **8**, 270–279 (2011). [doi:10.4161/rna.8.2.15013](https://doi.org/10.4161/rna.8.2.15013) [Medline](#)
 11. E. Minskaia, T. Hertzog, A. E. Gorbalenya, V. Campanacci, C. Cambillau, B. Canard, J. Ziebuhr, Discovery of an RNA virus 3'→5' exoribonuclease that is critically involved in coronavirus RNA synthesis. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 5108–5113 (2006). [doi:10.1073/pnas.0508200103](https://doi.org/10.1073/pnas.0508200103) [Medline](#)
 12. V. A. Avanzato, M. J. Matson, S. N. Seifert, R. Pryce, B. N. Williamson, S. L. Anzick, K. Barbian, S. D. Judson, E. R. Fischer, C. Martens, T. A. Bowden, E. de Wit, F. X. Riedo, V. J. Munster, Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. *Cell* **183**, 1901–1912.e9 (2020). [doi:10.1016/j.cell.2020.10.049](https://doi.org/10.1016/j.cell.2020.10.049) [Medline](#)
 13. T. Aydiello, A. S. Gonzalez-Reiche, S. Aslam, A. van de Guchte, Z. Khan, A. Obla, J. Dutta, H. van Bakel, J. Aberg, A. Garcia-Sastre, G. Shah, T. Hohl, G. Papanicolaou, M.-A. Perales, K. Sepkowitz, N. E. Babady, M. Kamboj, Shedding of Viable SARS-CoV-2 after Immunosuppressive Therapy for Cancer. *N. Engl. J. Med.* **383**, 2586–2588 (2020). [doi:10.1056/NEJMc2031670](https://doi.org/10.1056/NEJMc2031670) [Medline](#)
 14. B. Choi, M. C. Choudhary, J. Regan, J. A. Sparks, R. F. Padera, X. Qiu, I. H. Solomon, H.-H. Kuo, J. Boucau, K. Bowman, U. D. Adhikari, M. L. Winkler, A. A. Mueller, T. Y.-T. Hsu, M. Desjardins, L. R. Baden, B. T. Chan, B. D. Walker, M. Lichterfeld, M. Brigl, D. S. Kwon, S. Kanjilal, E. T. Richardson, A. H. Jonsson, G. Alter, A. K. Barczak, W. P. Hanage, X. G. Yu, G. D. Gaiha, M. S. Seaman, M. Cernadas, J. Z. Li, Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* **383**, 2291–2293 (2020). [doi:10.1056/NEJMc2031364](https://doi.org/10.1056/NEJMc2031364) [Medline](#)
 15. M. K. Hensley *et al.*, Prolonged SARS-CoV-2 infection CART. OSF Preprints osf.io/hmhse (2020).
 16. Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017). [doi:10.2807/1560-7917.ES.2017.22.13.30494](https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494) [Medline](#)
 17. X. Chi, R. Yan, J. Zhang, G. Zhang, Y. Zhang, M. Hao, Z. Zhang, P. Fan, Y. Dong, Y. Yang, Z. Chen, Y. Guo, J. Zhang, Y. Li, X. Song, Y. Chen, L. Xia, L. Fu, L. Hou, J. Xu, C. Yu, J. Li, Q. Zhou, W. Chen, A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020). [doi:10.1126/science.abc6952](https://doi.org/10.1126/science.abc6952) [Medline](#)
 18. C. O. Barnes, A. P. West Jr., K. E. Huey-Tubman, M. A. G. Hoffmann, N. G. Sharaf, P. R. Hoffman, N. Koranda, H. B. Grinstead, C. Gaebler, F. Muecksch, J. C. C. Lorenzi, S. Finkin, T. Hägglöf, A. Hurley, K. G. Millard, Y. Weisblum, F. Schmidt, T. Hatziioannou, P. D. Bieniasz, M. Caskey, D. F. Robbiani, M. C. Nussenzweig, P. J. Bjorkman, Structures of Human Antibodies Bound to SARS-CoV-2 Spike Reveal Common Epitopes and Recurrent Features of Antibodies. *Cell* **182**, 828–842.e16 (2020). [doi:10.1016/j.cell.2020.06.025](https://doi.org/10.1016/j.cell.2020.06.025) [Medline](#)
 19. L. Liu, P. Wang, M. S. Nair, J. Yu, M. Rapp, Q. Wang, Y. Luo, J. F.-W. Chan, V. Sahi, A. Figueroa, X. V. Guo, G. Cerutti, J. Bimela, J. Gorman, T. Zhou, Z. Chen, K.-Y. Yuen, P. D. Kwong, J. G. Sodroski, M. T. Yin, Z. Sheng, Y. Huang, L. Shapiro, D. D. Ho, Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020). [doi:10.1038/s41586-020-2571-7](https://doi.org/10.1038/s41586-020-2571-7) [Medline](#)
 20. W. B. Klimstra, N. L. Tilston-Lunel, S. Nambulli, J. Boslett, C. M. McMillen, T. Gilliland, M. D. Dunn, C. Sun, S. E. Wheeler, A. Wells, A. L. Hartman, A. K. McElroy, D. S. Reed, L. J. Rennick, W. P. Duprex, SARS-CoV-2 growth, furin-cleavage-site adaptation and neutralization using serum from acutely infected hospitalized COVID-19 patients. *J. Gen. Virol.* **101**, 1156–1169 (2020). [Medline](#)
 21. K. S. Xue, T. Stevens-Ayers, A. P. Campbell, J. A. Englund, S. A. Pergam, M. Boeckh, J. D. Bloom, Parallel evolution of influenza across multiple spatiotemporal scales. *eLife* **6**, e26875 (2017). [doi:10.7554/eLife.26875](https://doi.org/10.7554/eLife.26875) [Medline](#)
 22. Y. Weisblum, F. Schmidt, F. Zhang, J. DaSilva, D. Poston, J. C. C. Lorenzi, F. Muecksch, M. Rutkowska, H.-H. Hoffmann, E. Michailidis, C. Gaebler, M. Agudelo, A. Cho, Z. Wang, A. Gazumyan, M. Cipolla, L. Luchsinger, C. D. Hillyer, M. Caskey, D. F. Robbiani, C. M. Rice, M. C. Nussenzweig, T. Hatziioannou, P. D. Bieniasz, Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *eLife* **9**, e61312 (2020). [doi:10.7554/eLife.61312](https://doi.org/10.7554/eLife.61312) [Medline](#)
 23. L. Piccoli, Y.-J. Park, M. A. Tortorici, N. Czudnochowski, A. C. Walls, M. Beltramello, C. Silacci-Fregni, D. Pinto, L. E. Rosen, J. E. Bowen, O. J. Acton, S. Jaconi, B. Guarino, A. Minola, F. Zatta, N. Sprugasci, J. Bassi, A. Peter, A. De Marco, J. C. Nix, F. Mele, S. Jovic, B. F. Rodriguez, S. V. Gupta, F. Jin, G. Piumatti, G. Lo Presti, A. F. Pellanda, M. Biggiogero, M. Tarkowski, M. S. Pizzuto, E. Camerini, C. Havenar-Daughton, M. Smithey, D. Hong, V. Lepori, E. Albanese, A. Ceschi, E. Bernasconi, L. Elzi, P. Ferrari, C. Garzoni, A. Riva, G. Snell, F. Sallusto, K. Fink, H. W. Virgin, A. Lanzavecchia, D. Corti, D. Veesler, Mapping Neutralizing and Immunodominant Sites on the SARS-CoV-2 Spike Receptor-Binding Domain by Structure-Guided High-Resolution Serology. *Cell* **183**, 1024–1042.e21 (2020). [doi:10.1016/j.cell.2020.09.037](https://doi.org/10.1016/j.cell.2020.09.037) [Medline](#)
 24. D. Li *et al.*, The functions of SARS-CoV-2 neutralizing and infection-enhancing antibodies in vitro and in mice and nonhuman primates. *bioRxiv* [424729](https://doi.org/10.1101/2021.01.02.424729) [preprint]. 2 January 2021.
 25. W. N. Voss *et al.*, Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes in COVID-19 convalescent plasma. *bioRxiv* [423708](https://doi.org/10.1101/2020.12.21.423708) [preprint]. 21 December 2020.
 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. Kempet *et al.*, Recurrent emergence and transmission of a SARS-CoV-2 Spike deletion ΔH69/ΔV70. *bioRxiv* [422555](https://doi.org/10.1101/2020.12.21.422555) [preprint]. 21 December 2020.
 28. H. Tegally *et al.*, Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa. *medRxiv* [248640](https://doi.org/10.1101/2020.12.22.428640) [preprint]. 22 December 2020.
 29. A. Watanabe, K. R. McCarthy, M. Kuraoka, A. G. Schmidt, Y. Adachi, T. Onodera, K. Tonouchi, T. M. Caradonna, G. Bajic, S. Song, C. E. McGee, G. D. Sempowski, F. Feng, P. Urick, T. B. Kepler, Y. Takahashi, S. C. Harrison, G. Kelsoe, Antibodies to a Conserved Influenza Head Interface Epitope Protect by an IgG Subtype-Dependent Mechanism. *Cell* **177**, 1124–1135.e16 (2019). [doi:10.1016/j.cell.2019.03.048](https://doi.org/10.1016/j.cell.2019.03.048) [Medline](#)
 30. K. Katoh, K. Misawa, K. Kuma, T. Miyata, MAFFT: A novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res.* **30**, 3059–3066 (2002). [doi:10.1093/nar/gkf436](https://doi.org/10.1093/nar/gkf436) [Medline](#)
 31. K. Katoh, D. M. Standley, MAFFT multiple sequence alignment software version 7: Improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013). [doi:10.1093/molbev/mst010](https://doi.org/10.1093/molbev/mst010) [Medline](#)
 32. M. N. Price, P. S. Dehal, A. P. Arkin, FastTree: Computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009). [doi:10.1093/molbev/msp077](https://doi.org/10.1093/molbev/msp077) [Medline](#)
 33. A. Stamatakis, RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* **30**, 1312–1313 (2014). [doi:10.1093/bioinformatics/btu033](https://doi.org/10.1093/bioinformatics/btu033) [Medline](#)

34. K. H. D. Crawford, R. Eguia, A. S. Dingens, A. N. Loes, K. D. Malone, C. R. Wolf, H. Y. Chu, M. A. Tortorici, D. Veessler, M. Murphy, D. Pettie, N. P. King, A. B. Balazs, J. D. Bloom, Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* **12**, 513 (2020). [doi:10.3390/v12050513](https://doi.org/10.3390/v12050513) [Medline](#)

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 Stephen C. Harrison for his support. We thank Dr. Alison Morris, Dr. Bryan McVerry, Dr. Georgios Kitsios, Dr. Barbara Methe, Heather Michael, Michelle Busch, John Ries, and Caitlin 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 CR19050099-009. **Funding:** This work was supported by The University of Pittsburgh, the Center for Vaccine Research, The Richard King Mellon Foundation, the Hillman Family Foundation (WPD) and UPMC Immune Transplant and Therapy Center (WGB, GH). **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. 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. **Data and materials 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.gsaaid.org; see table S2 for full list of Acknowledgments). This work is licensed under a Creative Commons Attribution 4.0 International (CC BY 4.0) license, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited. To view a copy of this license, visit <https://creativecommons.org/licenses/by/4.0/>. This license does not apply to figures/photos/artwork or other content included in the article that is credited to a third party; obtain authorization from the rights holder before using such material.

SUPPLEMENTARY MATERIALS

science.sciencemag.org/cgi/content/full/science.abf6950/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

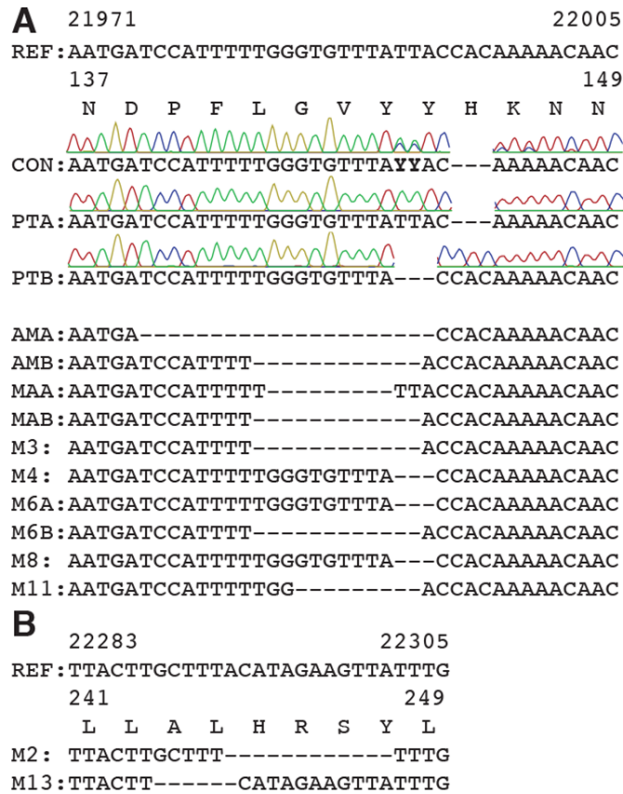


Fig. 1. Deletions in SARS-CoV-2 spike arise during persistent infections of immunosuppressed patients. (A) Top. Sequences of viruses isolated from PLT11 (PT) and viruses from patients with deletions in the same NTD region. Chromatograms are shown for sequences from PLT11, which include sequencing of 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 2, 3, 4, 6, 8, 11, 13 (13) Letters (A and B) designate different variants from the same patient. (B) Sequences of viruses from two patients with deletions in a different region of the NTD. All sequences are aligned to reference sequence (REF) MN985325 (WA-1). Genetic analysis of patient isolates is in fig. S1.

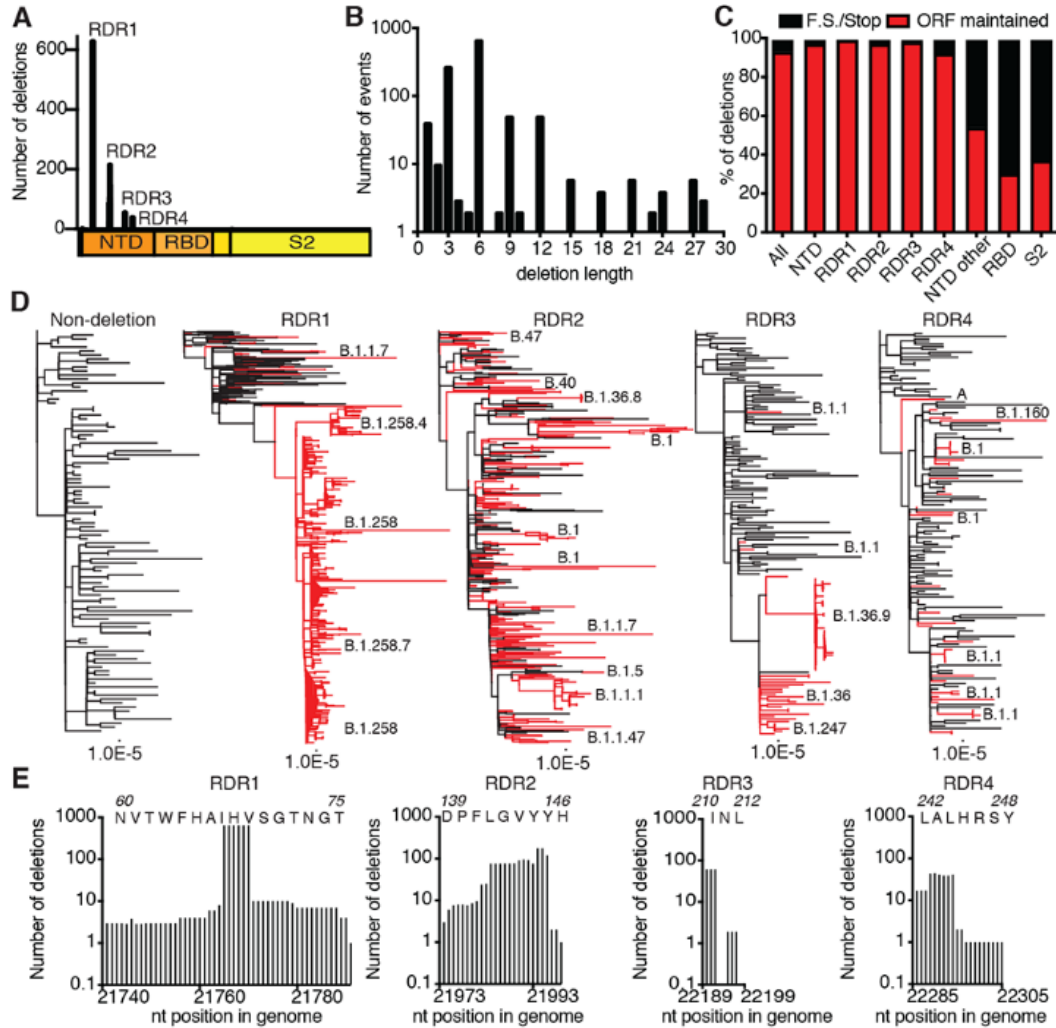


Fig. 2. Identification and characterization of recurrent deletion regions in SARS-CoV-2 spike protein. (A) Positional quantification of deleted nucleotides in S among GISAID sequences. We designate the four clusters recurrent deletion regions (RDRs) 1-4.). (B) Length distribution of deletions. (C) The percentage of deletion events at the indicated site that either maintain the open reading frame or introduce a frameshift or premature stop codon (F.S./Stop). (D) Phylogenetic analysis of deletion variants (red branches) and genetically diverse non-deletion 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 deletions in each RDR. Positions are defined by reference sequence MN985325, by codon (top) and nucleotide (below).

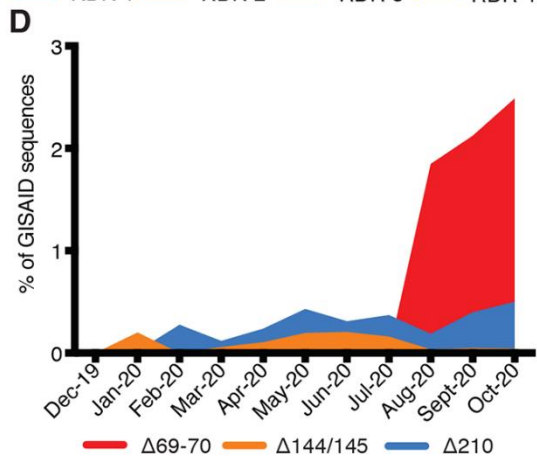
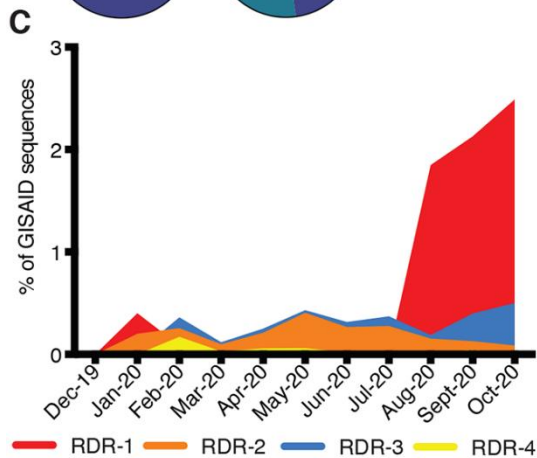
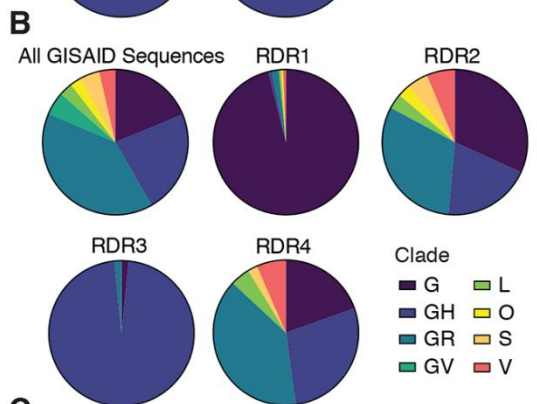
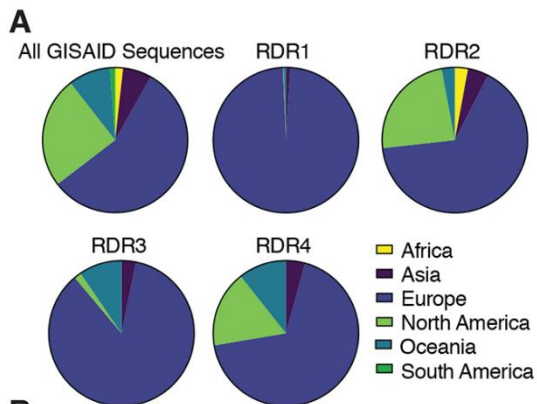


Fig. 3. Geographic, genetic, and temporal abundance of RDR variants. (A and B) Geographic (A) and genetic (B) distributions of RDR variants compared to the GISAID database (sequences from 12-1-2019 to 10-24-2020). GISAID clade classifications are used in (B). (C) Frequency of RDR variants among all complete 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/ Δ 210 has been adjusted by 0.02 units on the Y-axis for visualization in (C) due to its overlap with RDR2 and this adjustment has been retained in (D) to make direct comparisons between panels.

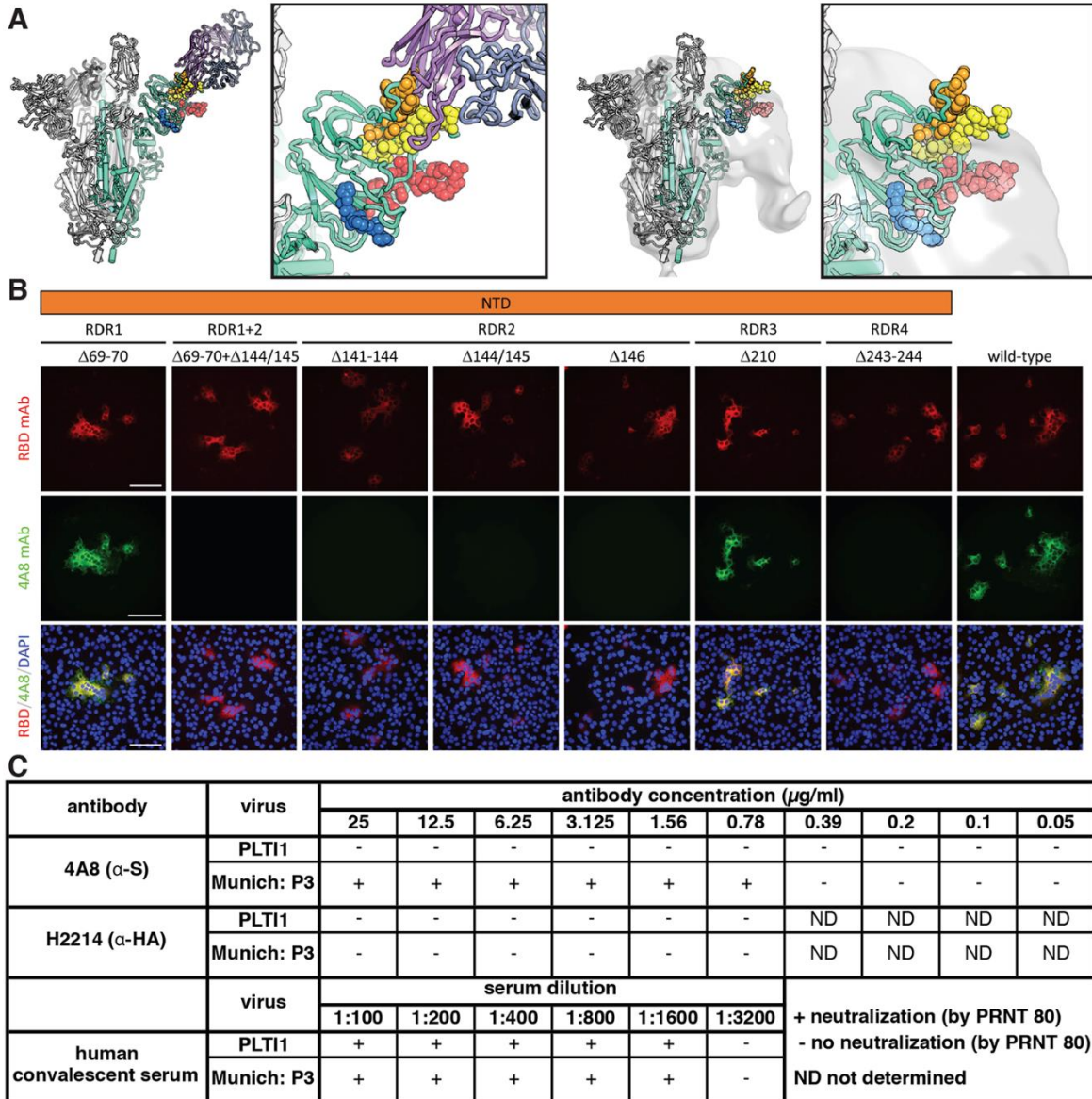


Fig. 4. Deletions in the spike NTD alter its antigenicity; RDRs map to defined antigenic sites. (A) Top: A structure of antibody 4A8 (17) (PDB: 7C21) (purples) bound to one protomer (green) of a SARS-CoV-2 spike trimer (grays). RDRs 1-4 are colored red, orange, blue, and yellow, respectively, and shown in spheres. The interaction site is shown at right. Bottom: The electron microscopy density of COV57 serum Fabs (18) (EMDB emd_22125) fit to SARS-CoV-2 S glycoprotein trimer (PDB: 7C21). The same view of the interaction site is provided at right. (B) S glycoprotein distribution in Vero E6 cells at 24 h post-transfection with S protein deletion mutants, visualized by indirect immunofluorescence in permeabilized cells. A monoclonal antibody against SARS-CoV-2 S protein receptor-binding domain (RBD MAb; red) detects all mutant forms of the protein ($\Delta 69-70$, $\Delta 69-70+\Delta 141-144$, $\Delta 141-144$, $\Delta 144/145$, $\Delta 146$, $\Delta 210$ and $\Delta 243-244$) and the unmodified protein (wild-type). 4A8 monoclonal antibody (4A8 MAb; green) does not detect mutants containing deletions in RDR2 or RDR4 ($\Delta 69-70+\Delta 141-144$, $\Delta 141-144$, $\Delta 144/145$, $\Delta 146$ and $\Delta 243-244$). Overlay images (RBD/4A8/DAPI) depict co-localization of the antibodies; nuclei were counterstained with DAPI (blue). The scale bars represent 100 μm . (C) Virus isolated from PLT11 resists neutralization by 4A8. A non-deletion variant (Munich) is neutralized by 4A8, both are neutralized by convalescent serum and neither is neutralized by an influenza hemagglutinin binding antibody H2214 (29).

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

Kevin R. McCarthy, Linda J. Rennick, Sham Nambulli, Lindsey R. Robinson-McCarthy, William G. Bain, Ghady Haidar and W. Paul Duprex

published online February 3, 2021

ARTICLE TOOLS

<http://science.sciencemag.org/content/early/2021/02/02/science.abf6950>

SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2021/02/02/science.abf6950.DC1>

REFERENCES

This article cites 32 articles, 4 of which you can access for free
<http://science.sciencemag.org/content/early/2021/02/02/science.abf6950#BIBL>

PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

Science (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. The title *Science* is a registered trademark of AAAS.

Copyright © 2021 The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. Distributed under a Creative Commons Attribution License 4.0 (CC BY).