Tracking the mutation potential of SARS-CoV-2

By Dan Ryan

Introduction

"Mutation. It is the key to our evolution. It is how we have evolved from a single-cell organism into the dominant species on the planet." These prosaic words from the cerebrally enhanced but utterly fictitious Professor Charles Xavier imply a greater purpose. Mutations are just chance errors in the copying of DNA and RNA. Without them, generations would not be able to adapt to meet intense selective pressures, but the vast majority have no impact and a few are harmful, disrupting the way that proteins fold and interact. This is illustrated by the astounding accuracy of DNA copying in multi-cellular organisms. In human cells, we see an error rate of 1 in every 100,000,000 base pairs.

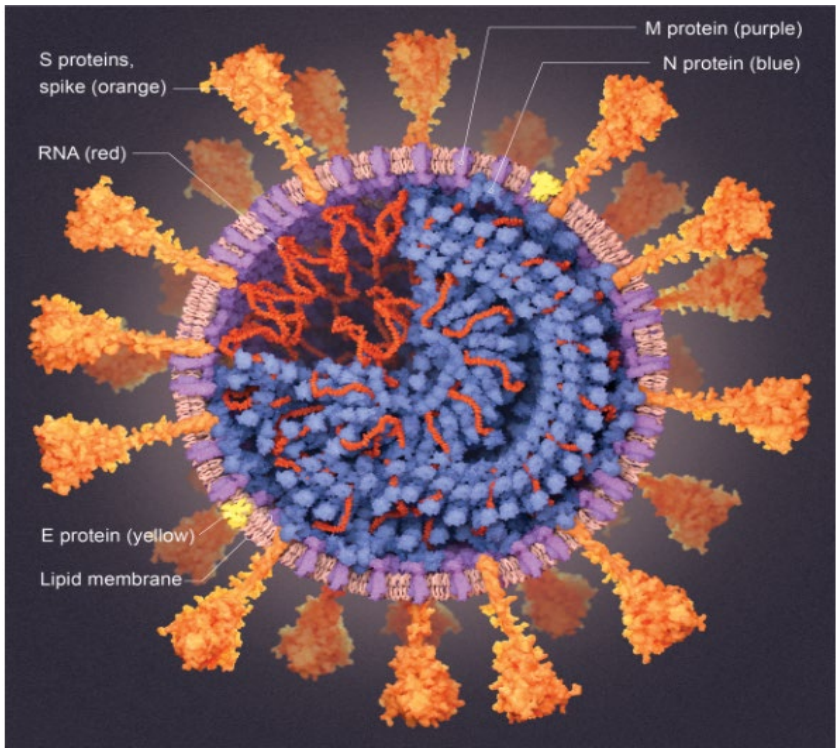
However, the more variable the environment, the greater the potential value of mutations and the greater the selective pressure. For viruses attacking a foreign host and withstanding the multiple layers of your immune defences, the challenge could not be more extreme. Generations of RNA viruses have been selected for faster replication over accuracy, and the result is much higher mutation rates, as high as 1 in every 1,000 base pairs[1]. That said,

Mutation rates of coronaviruses

The RNA genome of the SARS-CoV-2 virus is 30,000 base pairs long, approximately 100,000 times smaller than our own DNA. Nevertheless this is still quite a large genome for a virus, and so coronaviruses (unlike other RNA viruses) have developed a genetic proofreading mechanism to reduce the likelihood of errors in viral replication and hence are more stable with a somewhat lower rate of mutation[2].

The viral genome is much more efficient, coding for 4 major proteins as compared to 25,000 in humans. These major proteins include nucleocapsid (N), envelope (E), membrane (M) and the distinctive spike (S) protein[3].

Figure 1: Protein structure of SARS-CoV-2 virus (Credit: Veronica Falconieri Hays)

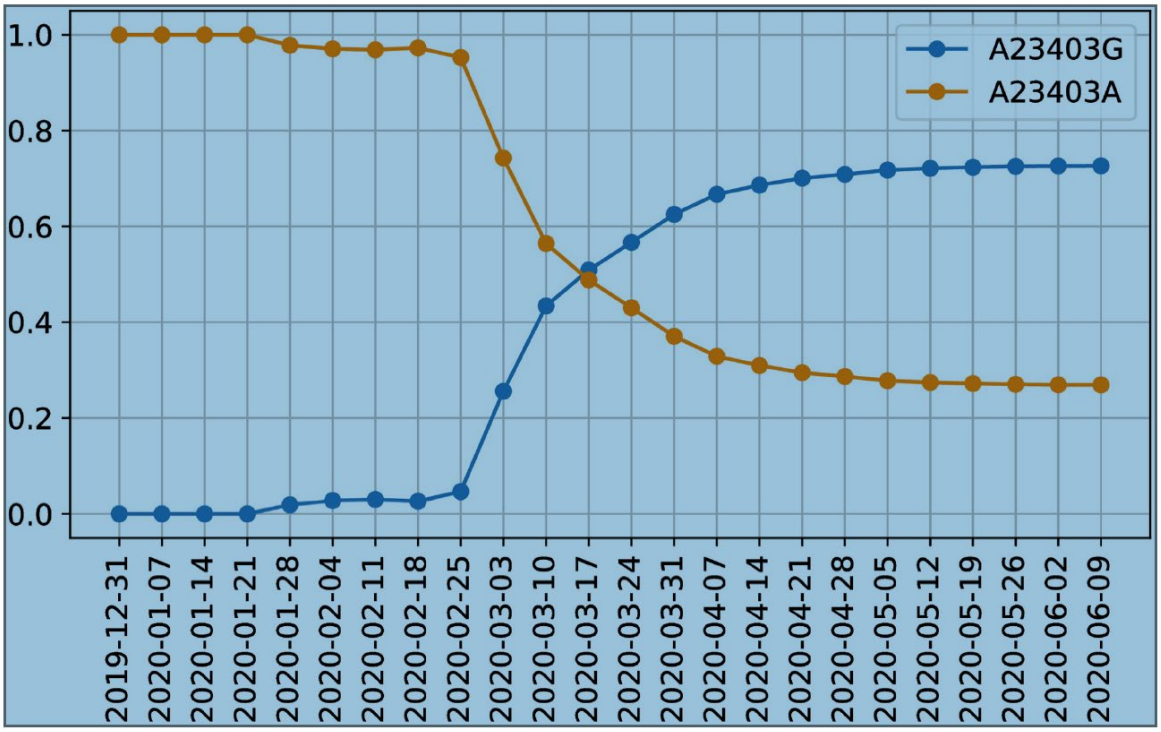


Our battle with COVID-19 has developed unprecedented insights into the foe that we face, with almost 150,000 samples sequenced and made publicly available in near real-time through collections such as the Global Initiative of Sharing All Influenza Data (GISAID). Nextstrain.org has developed a time map of viral distribution that tracks the emergence of new variants and determine the mutation rate of the virus. This appears to be 2 single base pair mutations per month, about half the rate of influenza and one quarter the rate of HIV. This analysis has been used to establish that the virus first infected humans back in November 2019[4].

The appearance of the D614G mutation

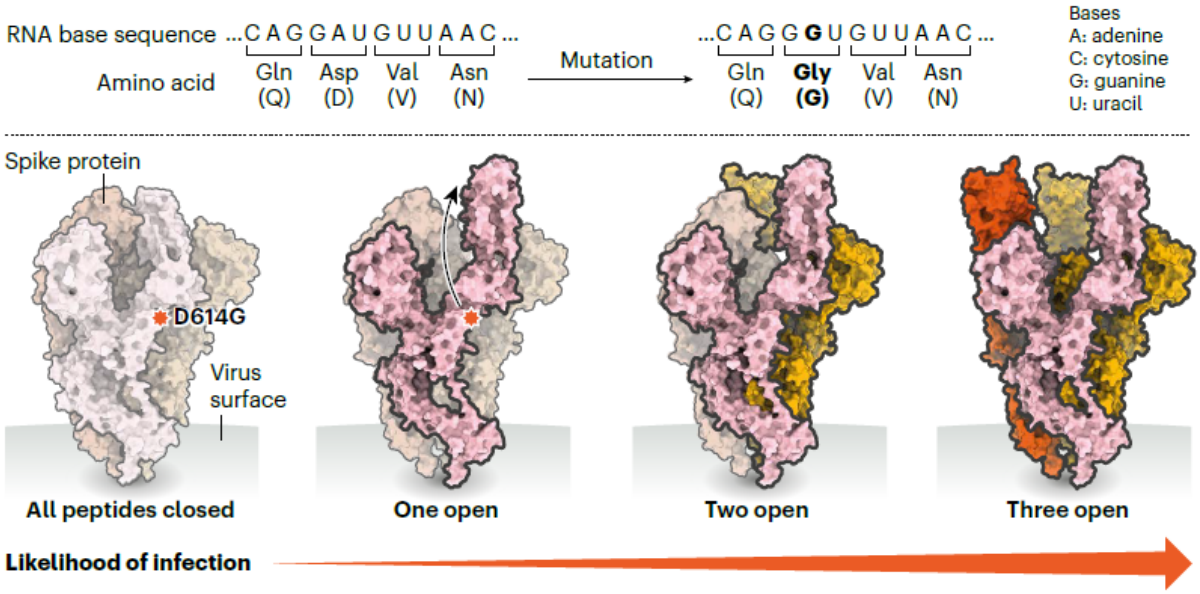
The initial L clade or variant from Wuhan, China was replaced by S and V variants as SARS-CoV-2 began to spread around the world. Analyses from late May 2020 started to detect a rapid increase in the prevalence of a D614G mutation that had rarely been seen before March 2020 but was the dominant variant by June 2020.

Figure 2: Relative abundance of variants of D614G spike protein mutation (A23403G)[5]



The mutation leads to an amino acid substitution (glycine (G) for aspartate (A)) at position 614 in the spike protein. The spike protein is composed of three peptides, and this substitution appears to "loosen" up the connections[6], making it more likely for infection to take place. Whilst this has now been confirmed through modelling and animal testing, it has not yet been confirmed clinically through tracking human clusters. The G variant replicates more efficiently in airway epithelial cells, develops higher concentrations of virus in nose and trachea and would appear to transmit earlier through droplets (in hamster trials).[7]

Figure 3. Illustration of impact of mutation on configuration of SARS-CoV-2 spike protein



All of the vaccines currently under testing are targeted on the distinctive spike protein. Antibodies are highly selective, and most vaccines were originally targeted on the original D614 form of the spike protein. Any mutation affecting the shape of the spike protein could reduce the effectiveness of a vaccine. However, animal and human sera tests indicate that the G614 variant is instead more susceptible to neutralisation and the effectiveness of current vaccines is not expected to be affected by the D614G mutation.[8]

Our vulnerability to further mutation

As of 12 October, there were 38,025,391 confirmed cases of COVID-19. Given the lack of testing particularly in the early months of the pandemic and the high percentage of asymptomatic cases, the true figure is much more likely to be north of 200 million. The database assembled by the London School of Hygiene and Tropical Medicine and in-depth studies of clusters suggests that the virus is dependent on a limited number of hosts for transmission. It has an unusually low dispersion factor, with perhaps 10% of initial infections being responsible for 80% of secondary transmissions[9]. Nevertheless, it has been successful and there is currently little selective pressure to promote further mutations. As we move into the winter months in the Northern Hemisphere, the numbers of positive tests are rising exponentially, indicating a sizeable susceptible population despite the first wave.

However, as we move into next year the advent of mass vaccination programmes may change the dynamic, and new mutations may prove to have a selective advantage. Our vaccination strategy against influenza is to predict which viral strains will be circulating in the winter season, targeting the various surface proteins of the virus. Our strategy so far with SARS-CoV-2 has been the same, focusing on the spike protein. In the same way that a Phase III trial[10] is testing whether an influenza vaccine can be developed to the less immunogenic but constant capsid proteins, perhaps some effort should be focused on developing vaccine alternatives for SARS-CoV-2 that similarly focus on the more constant nucleocapsid proteins.

References

1. Duffy S. Why are RNA virus mutation rates so damn high? *PLoS Biol*. 2018;16(8):e3000003. doi:10.1371/journal.pbio.3000003
2. Plante JA, Liu Y, Liu J, et al. *Spike Mutation D614G Alters SARS-CoV-2 Fitness and Neutralization Susceptibility*. Microbiology; 2020. doi:10.1101/2020.09.01.278689
3. Hoque MN, Chaudhury A, Akanda MAM, Hossain MA, Islam MT. Genomic diversity and evolution, diagnosis, prevention, and therapeutics of the pandemic COVID-19 disease. *PeerJ*. 2020;8:e9689. doi:10.7717/peerj.9689

4. Duchene S, Featherstone L, Haritopoulou-Sinanidou M, Rambaut A, Lemey P, Baele G. *Temporal Signal and the Phylodynamic Threshold of SARS-CoV-2*. Evolutionary Biology; 2020. doi:10.1101/2020.05.04.077735
5. Zhao Z, Sokhansanj BA, Malhotra C, Zheng K, Rosen GL. Genetic grouping of SARS-CoV-2 coronavirus sequences using informative subtype markers for pandemic spread visualization. Segata N, ed. *PLoS Comput Biol*. 2020;16(9):e1008269. doi:10.1371/journal.pcbi.1008269
6. Mansbach RA, Chakraborty S, Nguyen K, Montefiori DC, Korber B, Gnanakaran S. *The SARS-CoV-2 Spike Variant D614G Favors an Open Conformational State*. Biophysics; 2020. doi:10.1101/2020.07.26.219741
7. Hou YJ, Chiba S, Halfmann P, et al. *SARS-CoV-2 D614G Variant Exhibits Enhanced Replication* Ex Vivo *and Earlier Transmission* in Vivo. Microbiology; 2020. doi:10.1101/2020.09.28.317685
8. Weissman D, Alameh M-G, de Silva T, et al. *D614G Spike Mutation Increases SARS CoV-2 Susceptibility to Neutralization.* Infectious Diseases (except HIV/AIDS); 2020. doi:10.1101/2020.07.22.20159905
9. Endo A, Centre for the Mathematical Modelling of Infectious Diseases COVID-19 Working Group, Abbott S, Kucharski AJ, Funk S. Estimating the overdispersion in COVID-19 transmission using outbreak sizes outside China. *Wellcome Open Res*. 2020;5:67. doi:10.12688/wellcomeopenres.15842.3
10. Phillipson J. Last of 12,400 Participants Completes Final Visit in BiondVax's M-001 Universal Flu Vaccine Pivotal Phase 3 Clinical Trial. BiondVax. Accessed October 13, 2020. https://www.biondvax.com/2020/07/last-of-12400-participants-completes-final-visit-in-biondvaxs-m-001-universal-flu-vaccine-pivotal-phase-3-clinical-trial/