A glimpse into the origins of genetic diversity in SARS-CoV-2

Joel O. Wertheim[1]

[1]Department of Medicine, University of California San Diego, San Diego, CA, 92103

Email: jwertheim@ucsd.edu

A deadly pathogen has found a new host. Near the end of 2019, the novel coronavirus SARS-CoV-2 began transmitting among humans. In the first two months following its discovery, over 75,000 cases of COVID-19 causing over 2,000 deaths have been confirmed. Researchers rapidly characterized the viral genome [1] and have already identified a closely related virus in bats [2], pointing to these animals as the likely natural reservoir of this virus. The search continues for an intermediate host animal that may have facilitated the zoonosis, like camels do for MERS-CoV and palm civets did for SARS-CoV.

From a clinical and epidemiological perspective, there is obvious interest in whether SARS-CoV-2 will adapt to become more transmissible and/or virulent in humans. Before we can address this question, it is importantly to realize that most evolutionary options available to this virus have already been explored in bats, where there is an immense array of genetically distinct coronaviruses. Viruses in the betacoronavirus genus, to which SARS-CoV-2 belongs, have likely been infecting bats for tens of millions of years, possibly since the origin of bats themselves [3]. Within this natural reservoir, viruses exist that would cause high and low virulence in humans, as well as those viruses that are easily transmissible from human to human. Evolution tinkers with these viruses in bats, and the epidemiological consequences are seen both in pathogenic zoonotic diseases (e.g., SARS, MERS, and COVID-19) and in the less-virulent circulating coronaviruses causing common colds. Occasionally, however, some evolutionary avenues are not available to a zoonotic pathogen until after arrival in humans [4].

Evolution requires genetic diversity, and genetic diversity arises from two distinct mechanisms: mutation and recombination. Mutation is the replacement of one nucleotide base with another, or the addition/removal of nucleotide(s). Mutations are the only true source of genetic novelty. Recombination, on the other hand, can increase genetic diversity by placing existing mutations into new genetic backgrounds. Two articles in the most recent issue of *Clinical Infectious Diseases* explore these sources of genetic variation in SARS-CoV-2. Li and colleagues describe the genetic variability within 7 COVID-19 patients from China. Yi explores potential instances of recombination among 84 SARS-CoV-2 genomes from multiple countries.

Molecular epidemiology can use the genetic variation of SARS-CoV-2 to trace its history and better understand clusters of transmission. When we examine viruses infecting two different people, the amount of genetic similarity between these viruses is a proxy for the directness of their epidemiological relationship. Direct transmission partners are more likely to have identical or nearly identical sequences. Bear in mind that the genetic differences separating these viruses did not arise between these individuals; rather, all mutations arise during intra-host replication.

In this issue, Li and colleagues reveal the intra-host genetic variation of SARS-CoV-2. When these researchers examined genomes from a probable transmission pair who resided in the same house, they found the consensus sequences to be identical. However, deep-sequencing of viral

genomes from 7 patients (including this probable transmission pair), the authors detect substantial variation in the number and frequency of minority variants within different individuals. Remarkably, despite the limited number of individuals included, they found many of the same minority variants in different, epidemiologically unrelated individuals. This pattern suggests that SARS-CoV-2 is predisposed towards particular mutations, which may continue to re-appear as the virus continues to spread.

Recombination is an ongoing process in positive-strand RNA viruses [5], and coronaviruses are no exception. The evolutionary history of MERS-CoV in its intermediate host, camels, is littered with recombination events [6]. The bat coronaviruses from which SARS-CoV-2 descends also betray a history of recombination [1], although there does not appear to have been recombination distinguishing SARS-CoV-2 from its closest described relative in bats [7]. In contrast, the phylogeny for the first SARS-CoV, which emerged in 2002, does not display evidence for recombination. It is unlikely that this virus lacked the capacity for recombination, only the opportunity.

SARS-CoV-2 is presumably constantly recombining with itself within every infected individual. However, a homogenous population of viruses recombining with nearly identical relatives will not produce any novel genetic combinations. Genetic recombination has biological and evolutionary consequences only when the two recombining viruses are sufficiently genetically distinct.

In Yi's article, the prospect of recombination between genetically distinct SARS-CoV-2 strains is investigated. Robust phylogenetic testing for recombination cannot be performed on such closely related viruses, so Yi relies on detecting cycles in haplotype maps. These cycles represent potentially non-linear evolutionary histories of viral evolution that is consistent with recombination. However, among such closely related strains, the signal for recombination cannot be reliably disentangled from convergent evolution. A scenario for convergent evolution is plausible based the observation of recurrent minority variants reported by Li *et al*. Nonetheless, as more SARS-CoV-2 genomes are sequenced and analyzed, a clearer picture of the role recombination plays in SARS-CoV-2 will likely emerge.

The impact of this increasing genetic diversity on a hypothetical vaccine design is likely minimal. Vaccine efficacy endures against circulating human RNA viruses that have accumulated decades of evolutionary diversity (e.g., measles and polio virus). The obvious exceptions to this endurance are seasonal influenza viruses, where protective immunity extends over only a matter of years of viral evolution. Whether a successful vaccine is developed or SARS-CoV-2 continues to spread remains to be seen.

At the beginning of a zoonotic event, viral adaptation is difficult. The viral effective population size (an amalgam of the number of productive infections and the genetic diversity) is low; thus, natural selection will be inefficient, and recombination will not result in particularly novel combinations. However, as the number of infections increases and the circulating viruses become more genetically distinct, natural selection will become more efficient, making viral adaptation more of a possibility. Nonetheless, when SARS-CoV-2 arrived in humans, it was already well suited for person-to-person transmission, with pathogenic consequences. Furthermore, it is worth remembering that the evolutionary landscape that SARS-CoV-2 will explore in humans is dwarfed by its previous exploration in bats. For now, SARS-CoV-2 genetic variation is likely evolutionarily inconsequential and will be more important for facilitating molecular epidemiology in tracing the origins of novel clusters of viral infection.

Dr. Wertheim reports grants from National Institutes of Health (NIH), and grants from Centers for Disease Control and Prevention (CDC), outside the submitted work; He has no conflicts of interest to disclose.

REFERENCES

1. Lu R, Zhao X, Li J, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet **2020**.
2. Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature **2020**.
3. Wertheim JO, Chu DK, Peiris JS, Kosakovsky Pond SL, Poon LL. A case for the ancient origin of coronaviruses. J Virol **2013**; 87(12): 7039-45.
4. Pepin KM, Lass S, Pulliam JR, Read AF, Lloyd-Smith JO. Identifying genetic markers of adaptation for surveillance of viral host jumps. Nat Rev Microbiol **2010**; 8(11): 802-13.
5. Worobey M, Holmes EC. Evolutionary aspects of recombination in RNA viruses. J Gen Virol **1999**; 80 (Pt 10): 2535-43.
6. Dudas G, Carvalho LM, Rambaut A, Bedford T. MERS-CoV spillover at the camel-human interface. Elife **2018**; 7.
7. Paraskevis D, Kostaki EG, Magiorkinis G, Panayiotakopoulos G, Sourvinos G, Tsiodras S. Full-genome evolutionary analysis of the novel corona virus (2019-nCoV) rejects the hypothesis of emergence as a result of a recent recombination event. Infect Genet Evol **2020**; 79: 104212.