Variation and multilevel selection of SARS-CoV-2

Neil W Blackstone,1,2 Sarah R Blackstone,3 and Anne T Berg4

1Department of Biological Sciences, Northern Illinois University, DeKalb, Illinois 60115
2E-mail: neilb@niu.edu
3Department of Health Sciences, James Madison University, Harrisonburg, Virginia 22807
4Epilepsy Center, Ann & Robert H. Lurie Children’s Hospital of Chicago, Chicago, Illinois 60611

Received May 8, 2020
Accepted August 10, 2020

The evolution of SARS-CoV-2 remains poorly understood. Theory predicts a group-structured population with selection acting principally at two levels: the pathogen individuals and the group of pathogens within a single host individual. Rapid replication of individual viruses is selected for, but if this replication debilitates the host before transmission occurs, the entire group of viruses in that host may perish. Thus, rapid transmission can favor more pathogenic strains, while slower transmission can favor less pathogenic strains. Available data suggest that SARS-CoV-2 may follow this pattern. Indeed, high population density and other circumstances that favor rapid transmission may also favor more deadly strains. Health care workers, exposed to pathogenic strains of hospitalized patients, may be at greater risk. The low case fatality rate on the Diamond Princess cruise ship may reflect the founder effect—an initial infection with a mild strain. A vaccine made with one strain may confer limited immunity to other strains. Variation among strains may lead to the rapid evolution of resistance to therapeutics. Finally, if less pathogenic strains are largely associated with mild disease, rather than treating all SARS-CoV-2 positive individuals equally, priority could be focused on testing and contact tracing the most seriously symptomatic patients.

KEY WORDS: Coronavirus, epidemiology, levels of selection, transmission dynamics.

Severe acute respiratory syndrome (SARS) coronavirus (SARS-CoV) triggered a major human pandemic in 2002–2003 (Cheng et al. 2007). Similar beta-coronaviruses are endemic to horseshoe bat populations (Hu et al. 2017), and another outbreak appeared to be inevitable at the time (Cheng et al. 2007). Indeed, with the emergence of SARS-CoV-2, a second pandemic, coronavirus disease 2019 (COVID-19), is now well underway. SARS-CoV and SARS-CoV-2 show about 80% similarity in their genomes (Lu et al. 2020), corresponding to substantial diversity of these viruses in bat populations (Hu et al. 2017; Cui et al. 2019; Lu et al. 2020). Despite this background, little attention has been given to the potential for SARS-CoV-2 to evolve in human populations and the implications of such evolution for public health measures. Based on multilevel selection theory, we outline the potential course of evolution of SARS-CoV-2, preliminary data supporting this hypothesis, and the relevance of this knowledge to predictive models, public health practice, and policies.

A Multilevel Perspective

Since the time of Darwin, biologists have recognized that selection can operate at more than one level (Wilson and Wilson 2007). The multilevel view is often conceptualized as “individuals” (which may be multicellular organisms) and “groups of individuals.” Selection can favor traits that confer fitness at the individual level, or at the group level, or both. Conflicts between levels of selection—when selection favors a different trait at the lower level than at the higher—can also occur. Hypotheses that many traits evolved for the “good of the group” (e.g., Wynne-Edwards 1962) led to a careful examination of these evolutionary conflicts. Indeed, a strong consensus emerged that individual-level selection will overwhelm group selection under most circumstances (Williams 1966). When genetic relatives comprise groups, cooperation could be favored, but such kin selection was viewed as distinct from group selection (Hamilton 1964).
Further considerations of individual and group selection were stimulated by the Price equation (Price 1972), which describes selection at multiple hierarchical levels and “provides an ideal framework for addressing philosophical questions about the levels of selection” (Okasha 2006). Based on the Price equation, Shelton and Michod (2020) suggest that several types of selection should be considered in a group-structured population (e.g., within-group individual selection, between-group selection, and global individual selection). When a pathogen infects a host population, selection on the pathogen can be conceptualized as within-host individual selection, between-host group selection, and global individual selection. If the host population itself exhibits group structure, an additional level, between-host-group selection, may also be important.

The history of life provides the greatest opportunity for the application of multilevel selection, a theme developed in several paradigm-shifting works (Buss 1987; Maynard Smith and Szathmáry 1995; Michod 1999). From this perspective biological “individuals” such as chromosomes, cells, and multicellular organisms were derived by the banding together of lower-level units. Selection on these groups of lower-level units, which may have been kin, emerged during evolutionary transitions in individuality (Szathmáry 2015). Further, much of modern biology may reflect ancient mechanisms of conflict mediation that facilitated these transitions (Blackstone 1995; Radzvilavicius and Blackstone 2018). Thus, the question is not whether between-group selection can overcome within-group individual selection, but when and under what circumstances.

A structured population provides circumstances that favor between-group selection, as suggested in the epigraph (Luo 2014) for a viral pathogen. In a single, large population of pathogens, for instance those within a single host, the fastest replicator(s) will inexorably come to predominate if replication rate is heritable. Since pathogens typically infect multiple hosts, however, a group-structured population results, that is, a population divided into groups of pathogens within host individuals. These groups may be products of single infections and hence clonal except for mutational variation, or products of multiple infections, with greater diversity leading to greater within-host competition and the possibility of recombination. If transmission of the pathogen between hosts is rapid, population structure essentially disappears, and selection again favors the fastest replicator(s). In other words, rapid transmission diminishes between-host selection relative to within-host selection. However, if transmission is slow (for instance, because of host behavior), between-host selection on pathogens is potentiated. As described in the epigraph (Luo 2014), if a pathogen strain replicates too rapidly, transmission might not occur before the host is debilitated. Fast-replicating pathogen strains may thus face extinction. Slow-replicating strains that cause mild or inapparent disease may allow a longer window of transmission and persist in the host population.

Multilevel selection certainly does not rule out selection on viral traits other than replication and transmission, as discussed below. Likely, replication and transmission rates will be central to within- and between-group selection, but other traits (e.g., host switching) may be the target of global individual selection. Also, levels of selection above that of the host may be of consequence. Of particular relevance to the current pandemic, human hosts themselves exist in group-structured populations. Households, family gatherings, and other social groups play an outsized role in transmission of SARS-CoV-2 (e.g., Ghinai et al. 2020). Such host population structure complicates the use of simple statistics (e.g., reproductive number, R; Cobey 2020) to describe the pandemic. Calculations of R are most informative when hosts approximate a uniform distribution, and R values are unimodal. This is unlikely to be the case with SARS-CoV-2 (Fig. 1). Host population structure thus introduces a third level of selection—between host groups. In some cases (e.g., households, family gatherings), these may be kin groups, but in others (e.g., nursing homes), they clearly are not. While this is a very simple summary of what can be very complex evolutionary dynamics (e.g., Loverdo et al. 2012), it provides a useful starting point for evolutionary inference.

**Variation in SARS-CoV-2**

Without heritable variation, there can be no evolution. SARS-CoV-2 appears to be a remarkably stable retrovirus, with a mutation rate approximately $10^{-6}$ per nucleotide per replication cycle likely because of proof-reading mechanisms (Bar-on et al. 2020). Nevertheless, even with a low mutation rate, the extensive replication of this virus in human populations suggests that considerable variation has been produced. Indeed, some preliminary data...
support this hypothesis (Korber et al. 2020; Lorenzo-Redondo et al. 2020; Su et al. 2020; Tang et al. 2020). Focusing on the trimeric Spike protein, Korber et al. (2020) suggest that certain genetic variants or strains are rapidly and repeatedly replacing the initial SARS-CoV-2 strains worldwide. Possibly, there is variation in pathogenicity that responds to human containment measures or lack thereof (Tang et al. 2020). For instance, Lorenzo-Redondo et al. (2020) suggest that a Chicago strain has significantly lower viral loads than a New York strain, which may reflect the unusually rapid transmission that occurred in New York early in the pandemic (see below). Interestingly, some of these data suggest parallel evolution to that seen previously in SARS-CoV, involving nucleotide deletions that temper replication rate (Su et al. 2020). Earlier studies of the closely related SARS-CoV (Fig. 2) suggest that in animal hosts and in the initial human patients, orf8ab produced a single protein but that in some later human patients, a 29-nucleotide deletion occurred and divided the product into two proteins (Keng et al. 2006). One of these proteins appeared to downregulate replication, thus implying that as the pandemic proceeded, slower replication may have been favored under some conditions (Keng et al. 2006). Considerably more data on both variation in SARS-CoV-2 and the functional implications of this variation will no doubt be forthcoming. At this point, however, it appears that SARS-CoV-2 may be evolving as the multilevel theory predicts.

An Evolutionary Scenario

SARS-CoV-related coronaviruses (Fig. 2) have been evolving in bat and perhaps other animal populations for some time. Driven by global individual selection (Shelton and Michod 2020), a strain can jump to human hosts, and this has occurred to produce SARS-CoV and SARS-CoV-2. While to date the data are fragmentary and inconclusive, the following scenario remains plausible for the latter. During the initial outbreak, mutational variation accumulated. Rapid between-host transmission favored the strains that replicated the fastest, and these tended to be more pathogenic. Slower replicating and less pathogenic strains were present in the population but were not favored as long as the opportunity for between-host transmission in the population was high (e.g., crowding and other behavioral factors). Likely, a variety of strains spread globally. In human populations, the condition of the host influenced the effects of the strains, but there was a tendency for less pathogenic strains to cause mild or inapparent infections while more virulent strains caused more severe clinical disease. The policy response across the globe has been to institute mechanisms of diagnosis, isolation, and to limit person-to-person spread through social distancing. These conditions favored the less pathogenic, slower replicating strains. Nevertheless, under conditions that allowed rapid person-to-person transmission, the more pathogenic strains persisted and perhaps even proliferated.

This scenario is complicated by the structure of human populations noted above (Fig. 1). Human society is not a uniform dispersion of individuals, but a series of overlapping and interacting groups of individuals (e.g., households, family gatherings, nursing homes). SARS-CoV-2 spreads rapidly, almost inexorably, within these groups (e.g., Ghinai et al. 2020). Certainly, such rapid transmission will do little to hinder pathogenic strains. The key then to diminishing viral pathogenicity is slowing the transmission between these host groups. In terms of selection, these host groups may thus constitute the most relevant higher-level unit.

Predictions

Evolution is as close to a general theory of biology as there is. The value of any theory is that it allows predictions, which may or may not be supported by subsequent data. Based on selection
Rapid Transmission and Disease Severity

High levels of population density, multigenerational homes, reliance on mass transit, and other factors that are often related to poverty create the circumstances that facilitate rapid transmission by bringing large numbers of people into close, continuous contact. Multilevel selection suggests that this favors the most pathogenic strains and predicts more severe disease and a high case fatality in populations where crowding, for any reason, occurs. Given the available data, case fatality can be difficult to estimate due to inaccuracies in both numerators and denominators (Ioannidis 2020a, b). Nevertheless, areas such as New York City that exhibit a confluence of these circumstances also showed high population-based mortality rates early in the pandemic (Baker 2020). Conversely, as transmission rates slowed, less pathogenic strains may have been favored, and disease severity may have decreased. Controlling for confounding variables, Flacco et al. (2020) suggest that this occurred in Italy from March to April, although they attribute this effect to improved health care.

Hospital-Acquired Infections May Lead to Severe Cases

Anecdotal evidence suggests that health care workers may have a greater risk of developing severe illness despite their relatively younger age and better health than the general population (Glaser 2020). Pathogenic strains that evolve in areas subject to high transmission rates and cause severe illness will of course be concentrated in hospitals. Healthy health care workers that are routinely exposed to these pathogenic strains may thus be at greater risk for infection with strains that cause the most severe disease.

The Case of the Diamond Princess

Ioannidis (2020a, 2020b) points out that the cruise ship Diamond Princess provides an opportunity to determine with great precision case fatality ratio for COVID-19 (1%), which appears much lower than some other calculations would suggest particularly when adjusted for demographics. In evolutionary terms, this may be an example of the “founder effect,” in which the cruise ship was colonized by a particular strain or strains of SARS-CoV-2, which purely by chance happened to be of relatively low virulence. The result of this contained outbreak was a particularly low case fatality and may not necessarily be representative of mortality associated with other SARS-CoV-2 strains that are affecting other areas around the world.

Stealth Outbreaks of SARS-CoV-Related Coronaviruses

Some Asian nations have remarkably low numbers of SARS-CoV-2 cases. In part, this reflects public health measures developed during the SARS pandemic (Martin and Khan 2020). Possibly, other small outbreaks of SARS-CoV-related coronaviruses (Fig. 2) may have occurred without detection because the symptoms closely resemble influenza. Given the abundance and diversity of these viruses in Asian bat populations (Cheng et al. 2007; Hu et al. 2017; Cui et al. 2019; Lu et al. 2020), such outbreaks seem inevitable, and some may even have spread worldwide. Stealth outbreaks may have provided partial immunity in human populations, thus attenuating the clinical response to the SARS-CoV-2 in some cases, depending on the strain of the virus. Indeed, SARS-CoV-2-reactive CD4+ T cells were detected in 40–60% of unexposed individuals (Grifoni et al. 2020), although the authors attribute this to “common cold” coronaviruses.

Vaccine Development

It has long been recognized that immunity to one strain of a pathogen does not confer complete immunity to all strains. For instance, post-licensure testing of the early pneumococcal vaccines demonstrated lack of efficacy against serotypes not represented in the vaccine (Shapiro et al. 1991). The rapidly evolving influenza viruses require that a newly formulated vaccine be developed each year to protect against the most transmittable and most virulent strains present (e.g., Cox and Subbarao 1999). Ultimately, SARS-CoV-2 may behave similarly (Chumakov et al. 2020).

Resistance to Therapeutics

Remdesivir, a promising therapeutic, interferes with the action of the viral RNA polymerase by acting as a base-pair analog (Yin et al. 2020). Its action is reminiscent of azidothymidine (AZT), one of the first therapeutics in the HIV pandemic. Strain-to-strain variation in HIV in the target reverse transcriptase enzyme led to the rapid evolution of resistance to AZT (e.g., Herron and Freeman 2014). If similar variation exists in RNA polymerase in strains of SARS-CoV-2, rapid resistance may again result. The evolution of resistance should be considered as therapeutics are being developed. Stimulation of the innate immune response (Chumakov et al. 2020) may have advantages in this context.

Case Detection and Contact Tracing

The focus on testing mildly symptomatic or asymptomatic individuals is based on the premise that all variation in disease severity is due to characteristics of the host and not to differences in the specific strain of the SARS-CoV-2 virus. If, however, as suggested by the multilevel theory variation in clinical disease severity is driven to a significant degree by variation in the...
SARS-CoV-2 virus strains, taking a one-size-fits-all approach to all patients testing positive for a strain of the virus will provide uncertain gain at considerable cost. Testing and tracing contacts of individuals with severe disease would be most likely to identify those at greatest risk of contracting a severe infection. We cannot ignore the mild cases, however, as evolution can continue to occur. If mild strains are rapidly transmitted, they have the potential to evolve and become more pathogenic. In this context, it might be relevant that some less pathogenic strains may evolve by deletions (Keng et al. 2006; Su et al. 2020), which would be expected to have a very low rate of back-mutation. This might delay the re-emergence of pathogenicity in these mild strains.

In conclusion, it is imperative to consider the potential effects of evolution when considering a newly emerged and rapidly evolving pathogen. The SARS-CoV-2 is not a single entity. Evolutionary theory provides predictions to guide scientific investigations as well as public health policy. While considerable focus has been on characterizing co-morbidities of human hosts, equal emphasis should be given to characterizing the variation in the SARS-CoV-2. For example, are the SARS-CoV-2 genome sequences of infected individuals on the Diamond Princess similar or different from those that circulated in New York City early in the pandemic (cf., Korber et al. 2020; Lorenzo-Redondo et al. 2020)? Are more virulent strains now evolving in the explosive increase of cases in the southern and western United States? With accurate sequencing of the virus in individuals, epidemiological studies could determine the distribution of SARS-CoV-2 strains across regions of the world and within countries and link that information to infection rates and case fatality on a population level. On a more clinical level, the severity of COVID-19 presentation could be examined with respect to specific SARS-CoV-2 strains and tease apart what proportion of disease severity is related to the strain versus to the characteristics of the host. Evolutionary theory can guide our understanding of the SARS-CoV-2 pandemic. In doing so, it can provide the basis for rational and effective policy responses to minimize the hazards to health posed by COVID-19 while not making the response worse than the disease itself. Finally, we should keep in mind that the conditions that produced the SARS-CoV-2 pandemic have not changed. As human society occupies more and more of the biosphere, global individual selection (Shelton and Michod 2020) increasingly favors parasites that invade human-related niches—human crops, human livestock, and human beings.

AUTHOR CONTRIBUTIONS
All authors contributed to developing these ideas and writing the manuscript.

ACKNOWLEDGMENTS
Reviewer comments were very helpful for revising an earlier draft of this manuscript.

LITERATURE CITED
Baker, G. 2020. The unique catastrophe unfolding in New York City. Wall Street J. April 25–26:C2.
Bar-on, Y. M., A. Flamholz, R. Phillips, and R. Milo. 2020. SARS-CoV-2 (COVID-19) by the numbers. eLife 9:e57309.
Blackstone, N. W. 1995. A units-of-observation perspective on the endosymbiont theory of the origin of the mitochondrion. Evolution 49:785–796.
Buss, L. 1987. The evolution of individuality. Princeton Univ. Press, Princeton, NJ.
Cheng, V. C. C., S. K. P. Lau, P. C. Y. Woo, and K. Y. Yuen. 2007. Severe Acute Respiratory Syndrome coronavirus as an agent of emerging and reemerging infection. Clin. Microbiol. Rev. 20:660–694.
Chumakov, K. C., S. Benn, P. Aaby, S. Kottulien, and R. Gallo. 2020. Can existing live vaccines prevent COVID-19? Science 368:1187–1188.
Cobey, S. 2020. Modeling infectious disease dynamics. Science 368:713–714.
Cox, N. J., and K. Subbarao. 1999. Influenza. Lancet 354:1277–1282.
Cui, J., F. Li, and Z. L. Shi. 2019. Origin and evolution of pathogenic coronaviruses. Nat. Rev. Microbiol. 17:181–192.
Flacco, M. E., C. A. Martelucci, F. Bravi, G. Parruti, A. Mascitelli, L. Mantovani, and L. Manzoli. 2020. SARS-CoV-2 lethality decreased over time in two Italian Provinces. medRxiv. https://doi.org/10.1101/2020.05.23.20110882.
Ghini, I., S. Woods, K. A. Ritger, T. D. McPherson, S. R. Black, L. Sparrow, M. J. Fricchine, J. L. Kerins, M. Pacilli, P. S. Ruestow, et al. 2020. Community Transmission of SARS-CoV-2 at Two Family Gatherings—Chicago, Illinois, February–March 2020. MMWR 69:446–450.
Glaser, G. 2020. He ran marathons and was fit. So why did Covid-19 almost kill him? STAT News, April 21.
Grifoni, A., D. Weiskopf, S. I. Ramirez, J. Mateus, J. M. Dan, C. R. Moderbacher, S. A. Rawlings, A. Sutherland, L. Premkumar, R. S. Jadi, et al. 2020. Targets of t cell responses to SARS-CoV-2 Coronavirus in humans with COVID-19 Disease and unexposed individuals. Cell 181:1489–1501.
Hamilton, W. D. 1964. The genetical evolution of social behavior. I. J. Theor. Biol 7:1–16.
Herron, J. C., and S. Freeman. 2014. Evolutionary analysis. 5th ed. Pearson, New York, NY.
Hu, B., L. P. Zeng, X. L. Yang, X. Y. Ge, W. Zhang, B. Li, J. Z. Xie, X. R. Shen, Y. Z. Zhang, N. Wang, et al. 2017. Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. PLoS Pathogens 13:e1006698. https://doi.org/10.1371/journal.ppat.1006698.
Ioannidis, J. P. A. 2020a. A fiasco in the making? As the coronavirus pandemic takes hold, we are making decisions without reliable data. STAT News, March 17.
———. 2020b. Coronavirus disease 2019: the harms of exaggerated information and non-evidence-based measures. Eur. J. Clin. Invest 50:e13222.
Keng, C. T., Y. W. Choi, M. R. A. Wellokers, D. Z. L. Chan, S. Shen, S. G. Lim, W. Hong, and Y. J. Tan. 2006. The human severe acute respiratory syndrome virus (SARS-CoV) 8b protein is distinct from its counterpart in animal SARS-CoV and down regulates the expression of envelop protein in infected cells. Virology 354:132–142.
Korber, B., W. M. Fisher, S. Gnanakaran, H. Yoon, J. Theiler, W. Abfalterer, N. Hengartner, E. E. Giorgi, T. Bhattacharya, B. Foley, et al. 2020. Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell https://doi.org/10.1016/j.cell.2020.06.043.
Lam, T. T. Y., 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, et al. 2020. Identifying
