Fundamentals of genomic epidemiology, lessons learned from the coronavirus disease 2019 (COVID-19) pandemic, and new directions

Denis Jacob Machado PhD1, Richard Allen White III PhD1,2, Janice Kofsky PhD1 and Daniel A. Janies PhD2

1University of North Carolina at Charlotte, College of Computing and Informatics, Department of Bioinformatics and Genomics, Charlotte, North Carolina and
2University of North Carolina at Charlotte, North Carolina Research Campus (NCRC), Kannapolis, North Carolina

Abstract

The coronavirus disease 2019 (COVID-19) pandemic was one of the significant causes of death worldwide in 2020. The disease is caused by severe acute coronavirus syndrome (SARS) coronavirus 2 (SARS-CoV-2), an RNA virus of the subfamily Orthocoronavirinae related to 2 other clinically relevant coronaviruses, SARS-CoV and MERS-CoV. Like other coronaviruses and several other viruses, SARS-CoV-2 originated in bats. However, unlike other coronaviruses, SARS-CoV-2 resulted in a devastating pandemic. The SARS-CoV-2 pandemic rages on due to viral evolution that leads to more transmissible and immune evasive variants. Technology such as genomic sequencing has driven the shift from syndromic to molecular epidemiology and promises better understanding of variants. The COVID-19 pandemic has exposed critical impediments that must be addressed to develop the science of pandemics. Much of the progress is being applied in the developed world. However, barriers to the use of molecular epidemiology in low- and middle-income countries (LMICs) remain, including lack of logistics for equipment and reagents and lack of training in analysis. We review the molecular epidemiology literature to understand its origins from the SARS epidemic (2002–2003) through influenza events and the current COVID-19 pandemic. We advocate for improved genomic surveillance of SARS-CoV and understanding the pathogen diversity in potential zoonotic hosts. This work will require training in phylogenetic and high-performance computing to improve analyses of the origin and spread of pathogens. The overarching goals are to understand and abate zoonosis risk through interdisciplinary collaboration and lowering logistical barriers.

(Received 14 October 2021; accepted 15 October 2021)

How did genomic epidemiology become what it is?

Genomic epidemiology stems from molecular epidemiology, which uses evidence ranging from gel electrophoresis to multilocus sequence typing to study the origins and spread of pathogenic microorganisms. Janies et al1 reviewed the history of molecular epidemiology and compared it with syndromic epidemiology. Here, we focus on recent advances toward genomic epidemiology (Fig. 1), which includes genomic sequencing combined with rapid data sharing as enabled by the Internet. In 2002–2003, the severe acute respiratory syndrome coronavirus (SARS-CoV) was the first infectious disease for which scientists shared software and pathogen genetic data over the Internet to rapidly respond to the disease. Thereafter, genomic epidemiology was solidified by responses to H5N1, H1N1-2009, and other strains of influenza such as H7N92,3 and expanded to respond to foodborne and sexually transmitted diseases.2,4

The first SARS-CoV genome was shared after publication6,7 on National Center for Biotechnology Information’s (NCBI) GenBank website, which was customary. Meanwhile, dashboards, graphs, and maps emerged to track cases over time and space.8 Janies et al9,10 combined genomic and geographic data for SARS-CoV and H5N1 influenza, respectively, being the first to project phylogenies onto a virtual globe. Janies et al11 used Keyhole Markup Language (KML) to develop Supramap, which facilitates geographic mapping of phylogenies. Supramap allowed hypothesis testing ranging from the host and geographic origins of pathogens12 to tracing mutations that conferred drug resistance or host switching.13,14 Limitations of computing large data sets, coupled with a preference for sharing data after publication, resulted in a greater turnaround between data acquisition and results than occurs today. However, these conditions did not impede a hypothesis-driven field with value to decision makers, as demonstrated in a 2007 congressional hearing.15

In the 2000s, some genomes were sequenced for respiratory pathogens such as H1N1-2009. However, even SARS-CoV genomes were not always sequenced completely, and sequences were released gradually.9 This changed due to factors such as new DNA sequencing technologies.

How did advances in sequencing technology reshape genomic epidemiology?

Current genomic epidemiology of infectious diseases originated in response to the SARS-CoV epidemic.16 Sequencing the...
SARS-CoV genome was instrumental in recognizing it as a novel coronavirus associated with HCoV-OC43 and HCoV-229E.6,7 Researchers combined genomic and epidemiological data to trace the genotypic variation of the viral transmission paths between 2002 and 2003.17,18 However, today’s genomic surveillance evolved with the advance of high-throughput sequencing (HTS) (Fig. 1). Reuter et al19 summarized HTS history until 2015 and Pérez-Losada20 reviewed recent HTS advances. We focus on the sequence cost variation per raw megabase between 2001 and 202021 (Fig. 2a) to illustrate the increasing feasibility of sequencing coronavirus genomes (Fig. 2b). Considering raw nucleotide sequencing cost, US$100 was not sufficient to sequence one coronavirus genome in 2020, but $100 it would cover >400,000 genomes in 2020.

What are coronaviruses?

Coronaviruses correspond to the four genera of the subfamily Orthocoronavirinae. Gammacoronavirus (GammaCoVs) and Deltacoronavirus (DeltaCoVs) mainly infect birds and rarely infect mammals.22,23 Alphacoronavirus (AlphaCoVs) and Betacoronavirus (BetaCoVs) originated from Chiroptera (bats) and are often found in other mammals, including humans.24 The coronavirus virion encapsulates one of the longest RNA virus genomes (27–32 kb),25 which has complex gene expression26 and variable gene content among genera (Fig. 3a).27 Coronavirus infections in domestic animals are economically significant.28–30 However, the epidemic emergence of human coronaviruses (HCoVs) is a pressing concern because they cause infections in all age groups, often leading to respiratory or enteric diseases.31 Neurological illness or hepatitis is less frequent.32 The US Centers for Disease Control (CDC) website33 lists 7 HCoVs: 2 AlphaCoVs (HCoV-229E and HCoV-NL63) and 5 BetaCoVs (HCoV-OC43, HCoV-HKU1, SARS-CoV, MERS-CoV, and SARS-CoV-2). We added the human enteric coronavirus 4408 (HECV-4408) to the list because it was isolated from a child with acute gastroenteritis.34

How did SARS-CoV-2 accelerate the growth of genomic epidemiology?

Coronaviruses were not deemed highly pathogenic to humans until the 2002 SARS-CoV outbreak.35,36 The dangers of HCoVs were made more evident by the 2012 outbreak of Middle East respiratory syndrome (MERS) coronavirus (MERS-CoV).37 Nevertheless, coronaviruses did not receive the current level of attention until the pandemic coronavirus disease 2019 (COVID-19), caused by SARS-CoV-2, was first reported in humans in Wuhan, China, in December 2019.38 However, Pekar et al39 inferred that the virus was present in Hubei approximately a month before. On March 11, 2020, the World Health Organization (WHO) declared a pandemic due to the spread of SARS-CoV-2.38 By October 14, 2021, COVID-19 had caused 4,863,818 deaths worldwide.40 Understanding the emergence and evolution of SARS-CoV-2 is vital to preventing future pandemics.41 The question can be divided into 3 components. First, was the virus purposefully manipulated? Several peer-reviewed publications have concluded that SARS-CoV-2 emerged naturally via zoonosis (see eg, Anderson et al,42 Liu et al43, and Holmes et al44). Moreover, previous serology data indicate natural human infections by bat-hosted, SARS-like viruses.45 Second, was SARS-CoV-2 an accidental release? If a naturally occurring virus was transported to a laboratory and humans were infected shortly thereafter, the virus may not have accumulated sufficient mutations to record its passage through controlled environments.46 However, no evidence indicates that SARS-CoV-2 was known to scientists before December 2019.47,48 Third, what is the natural source of SARS-CoV-2? The most comprehensive phylogenetic analysis of coronavirus49 (Fig. 3b) addressed the fundamental evolution of HCoVs (Fig. 3c) and showed that SARS-CoV-2 results from bat-hosted viruses infecting humans.50 SARS-CoV-2 finds its closest related bat-hosted coronaviruses in the subgenus Sabecovirus, a subgroup of SARS-related coronaviruses (SARSr-CoV) first identified in horseshoe bats.
Bat-hosted viruses similar to SARS-CoV-2 were collected in the Yunnan province, >1,500 km away from Wuhan, but the hosts have a wide geographic range.\(^{45,52,53}\) Despite a confusing array of reports confirming\(^{54–56}\) and denying\(^{57}\) the origin of SARS-CoV-2 from pangolin (\textit{Manis javanica}) hosts, pangolins are not involved in the lineage of SARS-CoV-2 that infected humans.\(^{49}\) This finding is similar to the emergence of SARS-CoV,\(^{9}\) which also infected humans from bat-hosted viruses without any need for intermediate hosts, including Himalayan palm civets (\textit{Parguma larvata}) and raccoon dogs (\textit{Nyctereutes procyonoides}).

Are we sequencing SARS-CoV-2 genomes fast enough?

SARS-CoV-2 was identified on January 7, 2020. Three days later, its genome and metadata were shared via the Global Initiative on Sharing Avian Influenza Data (GISAID)\(^{58}\) EpiCoV database,\(^{59}\) before the first peer-reviewed article was published in February 2020.\(^{60}\)

To put the SARS-CoV-2 genome sequencing speed into context, consider that SARS-CoV was first reported in November 2002, but its genome was publicly released in April 2003.\(^{6}\) The speed at which such data are released was changed by several forces, illustrated by Janies et al.\(^{16}\) In brief, the reasons include the increased feasibility of genome sequencing, the willingness to share data before publication, and the rise of the popular GISAID database, which credits submitting laboratories.

Figure 4 shows the accumulation of 4,224,785 complete SARS-CoV-2 genomes in EpiCoV between January 10, 2020, and October 13, 2021. The curve is far from reaching a plateau, indicating that we are not producing coronavirus genomes at total capacity. Efforts to sequence SARS-CoV-2 following international guidelines\(^{61,62}\) are welcome because these data inform epidemiological forecasts (eg, increased transmission efficiency of SARS-CoV-2 variants has led to projections of the rise of higher numbers of cases\(^{63}\)).

Genomic sequencing generates a snapshot of a viral lineage in a place and time. When sequences are collected longitudinally, applications in genomic epidemiology and pandemic responses emerge, which we illustrate with 4 examples. First, profiling mutation fingerprints from the viral pangenome to individual infection quasi-species enables molecular contact tracing.\(^{64}\) Second, genomic sequencing informs the peptide mass fingerprinting (PMF) used to predict novel structures and find inhibitors for viral peptides,\(^{65}\) although results must be tested in randomized controlled trials\(^{66}\) to identify effective antivirals.\(^{67,68}\) Third, the data are used to model epidemic or pandemic size and severity.\(^{63}\) Fourth, viral sequences are fundamental for developing mRNA vaccines.\(^{69}\) For a review on current pitfalls and opportunities in applying HTS to SARS-CoV-2 genomes, see Chiara et al.\(^{70}\)

As SARS-CoV-2 becomes endemic,\(^{71,72}\) sequencing demand will remain high. SARS-CoV-2 infections are decreasing as more people develop immunity through natural infection or vaccination.\(^{73}\) However, variants may evade infection and vaccine-induced antibodies,\(^{74}\) especially with infections occurring months after vaccination (ie, breakthrough infections)\(^{75,76}\). Given breakthrough infections, increased transmission of some variants, and the lack of full vaccination among eligible people, we can predict that SARS-CoV-2 will continue to evolve. Whether SARS-CoV-2 is evolving toward more severe or more benign COVID-19 phenotypes is a pressing research question for genomic epidemiology.

Effective countermeasures depend on understanding SARS-CoV-2 lineages, such as sampling variants for which phenotype...
is not fully understood\(^77\) and addressing sampling bias.\(^78\) For example, if we restrict sequencing viral isolates from hospitalized patients, the relationships between any variables associated with hospitalization will be distorted when compared to the general population. Thus, we would miss mutations associated with asymptomatic and symptomatic cases that did not require hospitalization, which could lead to inducing or misinterpreting the evidence for phenotype-genotype associations.\(^79-81\)

Brito et al.\(^82\) analyzed the spatiotemporal heterogeneity in each country’s SARS-CoV-2 genomic surveillance efforts based on metadata submitted to GISAID until May 30, 2021. These researchers estimated that when the prevalence of a rare lineage is 2%, 300 cases would need to be sequenced to detect at least 1 genome of that lineage with 95% probability. Therefore, sequencing capacity should be at least 0.5% of cases per week when incidence is >100 positive cases per 100,000 people.

Brito et al.\(^82\) observed that countries like Denmark, which have a quick turnaround for sequencing, processing, and sharing SARS-CoV-2 genomic data (<18 days) and a high sequencing rate (>32%), observe greater lineage diversity. Many variants may be missed when sampling rates are low. However, disparities in wealth, investment in research and training, coordination, and supply chain logistics affect the ability of countries to perform genomic surveillance, especially LMICs. Therefore, efforts must be made to provide funds, training, and logistic support for researchers based in LMICs to improve their genomic surveillance capacity and public-health decision making.

**How do we classify the variants of SARS-CoV-2?**

Any genome sequence that is genetically distinct from the reference can be called a variant. In practice, the SARS-CoV-2 variants represent clades that share a set of key mutations while still permitting a small amount of other sequence variation.\(^83-84\) Moreover, convergent evolution among geographically distant variants has
been observed (Table 1). Although variants and strains are different, some researchers use these terms interchangeably (eg, Awadassie et al, Hossein et al, and Ul-Rahman et al). The term “strain” is typically associated with lineages that became sufficiently divergent to exhibit a changed phenotype.

In late 2020 and throughout 2021, as vaccine availability increased, information on variants began to dominate the COVID-19 response. The emergence of variants that might pose an increased risk to global public health prompted the WHO to characterize specific variants of interest (VOIs) and variants of concern (VOCs) to prioritize global monitoring and research. The US government SARS-CoV-2 Interagency Group (SIG) developed a separate variant classification scheme, which we compare to the WHO system in Table 2.

In March 2021, the WHO assigned letters of the Greek alphabet to categorize VOIs and VOCs for simplicity and to avoid association with particular localities. These labels do not replace existing classifications by GISAID, Nextstrain, and Pango lineages. SARS-CoV-2 variants were reviewed by Harvey et al.

### Why are vaccines still not enough against COVID-19?

The speed of development and testing of COVID-19 vaccines development is one of history’s most outstanding public health achievements. Vast vaccination of eligible individuals is the best and safest way to control the pandemic. Although some SARS-CoV-2 variants show a degree of escape from protective antibodies induced by natural infection (and, to a lesser degree, after immunization), T-cell responses are retained. Furthermore, first-generation SARS-CoV-2 mRNA-based vaccines induce public antibodies (ie, antibodies with similar genetic elements and modes of recognition against a different antigen observed in multiple individuals) with robust neutralizing and potentially durable protective activity against variants such as alpha (α), beta (β), and gamma (γ).

SARS-CoV-2 variants will continue to emerge, requiring close international monitoring to determine the need for vaccination boosters and or redesign. As variants emerge in areas of low vaccination, a global COVID-19 vaccination rollout is imperative. Since the vaccine rollout, new questions have arisen regarding vaccine efficacy against the transmission of different variants, the duration of protection, and the efficacy of prime-boost schedules. A demand has also arisen for studies to determine the immunological correlates of protection against COVID-19 as cases decline and prevention of severe disease gains more importance in vaccine efficacy. Meanwhile, nonpharmaceutical interventions to reduce the spread of SARS-CoV-2 and other pathogens are still warranted.

### How can we bridge the knowledge gap between disease origin and transmission?

Genomic epidemiology can be a tool to study emerging infectious diseases (EIDs) in humans, but its effectiveness is maximized when it accounts for animal and environmental components. In the case of zoonosis, there is a knowledge gap between the animal and human components of EID research, and One Health can bridge this gap.

Although most human health researchers have only started focusing on coronaviruses since the emergence of SARS-CoV-2, veterinarians, virologists, and zoologists have been researching animal coronaviruses long before the COVID-19 epidemic. One Health proposes placing these realms of research (on humans and animals) in the same environmental context. The next steps in pandemic prevention science are to understand factors that create opportunities for zoonosis, such as entering infectious habitats such as bat caves and the use of wildlife as food and medicine.

Deep sequencing the microbiomes and viromes of taxonomically, geographically, and temporally deep biorepository archives of putative host animals will serve as the basis of new approaches to zoonosis, risk assessment, and threat mitigation. Therefore, another step toward furthering the One Health approach is leveraging biorepositories in biomedical research. Although the Global Museum initiative already offers a route of international integration among museum biorepositories in a decentralized and geographically dispersed network, the link to EID research is still not fully realized.

---

**Table 1. Notable Variants of SARS-CoV-2 and Their Main Attributes**

| SIG Class | WHO Class | WHO Label | Pango Lineages | CDC Lineages | First Identified | Characteristic Spike Mutations |
|-----------|-----------|-----------|----------------|--------------|-----------------|--------------------------------|
| VOC VOC   | delta (δ) | Ay.1, Ay.2, B.1.617.2 | Ay             | India        | T19R, K417N, L452R, T478K, P681R, D950N |
| VBM VOC   | alpha (α) | B.1.1.7 and Q lineages | Q              | UK           | HV69-, Y144-, N501Y, A570D, P681H, T716I, S982A, D1118H |
| VBM VOC   | beta (β)  | B.1.351   | South Africa   | D89A, D215G, E484K, N501Y, A701V |
| VBM VOC   | gamma (γ) | P.1       | Brazil         | T20N, P625S, K417T, E484K, N501Y, T1027I |
| VBM VOI   | mu (μ)    | B.1.621 and B.1.621.1 | India         | T95I, Y145N, R346K, E484K, N501Y, D614G, P681H, D950N |
| N/A VOI   | lambda (λ) | C.37 | South America | G75V, T76I, R53LPG246-, L452Q, F490S, D614G, T859N |
| VBM VUM   | epsilon (ε) | B.1.427 and B.1.429 | USA           | S13I, W152C, L452R, D614G |
| VBM VUM   | eta (η)   | B.1.525   | UK and Nigeria | Q52R, E484K, G677H, F88L |
| VBM VUM   | iota (ι)  | B.1.526   | USA           | L5F, T95I, D215G, E484K, D614G, A701V, S477N, Q957R |
| VBM VUM   | kappa (κ) | B.1.617.1 | India         | L452R, E484Q, P681R |
| VBM VUM   | zeta (ζ)  | P.2       | Brazil         | E484K |
| VBM N/A   | N/A       | B.1.617.3 | India         | T19R, L452R, E484Q, P681R, D950N |

Note. SIG, US government SARS-CoV-2 Interagency Group; VBM, variant being monitored; VOC, variant of concern; VOI, variant of interest; VUM, variants under monitoring; EUA, emergency use authorization.

This table was modified and updated from the WHO website, the CDC website, Rambaut et al, and Soh et al. SIG and WHO classifications are detailed in Table 2.
The overarching goal of the MEPA is to guarantee rapid access to large volumes of raw and cynomolgus and rhesus macaques, and white-tailed ferrets, raccoons, and Syrian hamsters, and data standards. However, the likelihood of new mutations increases as cases rise, possibly leading to enhanced transmission, immune escape, diagnostic or therapeutic escape and identified to cause significant community transmission or multiple COVID-19 clusters, in multiple countries with increasing relative prevalence alongside increasing number of cases over time, or other apparent epidemiological impacts to suggest an emerging risk to global public health.

The need for this host-based approach to genomic epidemiology is made evident by the transmissible nature of SARS-CoV-2, which has the potential to infect a range of hosts, including tigers, raccoon dogs, Egyptian fruit bats, raccoon dogs, “taxa with exemplar approaches. Global efforts to strengthen pathogen sequencing capacity are designed to manage big data from rapidly evolving pathogens, like India, are not sequencing isolates at full capacity. The United States is an even more extreme example because it has ranked low in SARS-CoV-2 sequencing despite its capacity and expertise. On the other hand, countries like South Africa have sequencing laboratories struggling with reagent shortages and the scarcity of trained scientists.

Global efforts to strengthen pathogen sequencing capacity are still required to respond to technical, logistical, and financial challenges in resource-limited settings despite increased sequencing feasibility. Moreover, good SARS-CoV-2 sequencing performance for some LMICs (eg, Democratic Republic of the Congo, Brazil, Senegal, and Thailand) further encourages international and domestic collaboration among public health authorities, healthcare facilities, academia, and industries.

Additional challenges include consistent handling of isolates as well as metadata and sequence data curation and deposition in a way that facilitates combining data sets from different laboratories. These challenges require coordinated efforts and data standards to guarantee rapid access to large volumes of raw and processed molecular data at unprecedented scales.

How can we track SARS-CoV-2 variants faster?

Vaccines are still effective in preventing severe outcomes against all SARS-CoV-2 variants, which are ravaging unvaccinated people. However, the likelihood of new mutations increases as cases rise, possibly leading to enhanced transmission, immune escape, or increased pathogenicity. This process has resulted in more transmissible variants.

Researchers face 2 main challenges in keeping pace with SARS-CoV-2 variants: using resources at optimal capacity and lowering barriers to technology and training in genomic epidemiology across the world. On the one hand, countries with a high positivity rate, like India, are not sequencing isolates at full capacity.

| Table 2. Comparing the Different Categories in the WHO Variant Classification System with the System Used by the US government SARS-CoV-2 Interagency Group (SIG) |

| Category | Potential Attributes | Category | Working Definition |
|----------|---------------------|----------|--------------------|
| VBM      | Variants for which data indicate a potential or clear impact on approved or authorized medical countermeasures | VUM      | A SARS-CoV-2 variant with genetic changes that are suspected to affect virus characteristics with some indication that it may pose a future risk, but evidence of phenotypic or epidemiological impact is currently unclear, requiring enhanced monitoring and repeat assessment pending new evidence. |
| VOI      | Specific genetic markers that are predicted to affect transmission, diagnostics, or immune response | VOI      | With genetic changes that are predicted or known to affect virus characteristics such as transmissibility, disease severity, immune escape, diagnostic or therapeutic escape and identified to cause significant community transmission or multiple COVID-19 clusters, in multiple countries with increasing relative prevalence alongside increasing number of cases over time, or other apparent epidemiological impacts to suggest an emerging risk to global public health. |
| VOC      | Evidence of impact on diagnostics, treatments, or vaccines, evidence of increased transmissibility, evidence of increased disease severity, evidence of increased transmissibility, evidence of increased disease severity | VOC      | Increase in transmissibility or detrimental change in COVID-19 epidemiology OR increase in virulence or change in clinical disease presentation OR decrease in effectiveness of public health and social measures or available diagnostics, vaccines, therapeutics. |
| VOHC     | Impact on medical countermeasures, demonstrated failure of diagnostic test targets, evidence to suggest a significant reduction in vaccine effectiveness, a disproportionately high number of infections in vaccinated persons, very low vaccine-induced protection against severe disease, significantly reduced susceptibility to multiple EUA or approved therapeutics, more severe clinical disease and increased hospitalizations | VOHC     | (No equivalent WHO category) |

Note. VBM, variant being monitored; VOC, variant of concern; VOI, variant of interest; VUM, variants under monitoring; VOHC, variant of high consequence; EUA, emergency use authorization.

*Currently, no variants are being classified as VOI or VOHC by the CDC and SIG.
example, the ultrafast sample placement on existing trees (USher) enables the rapid placement of novel genomes into a reference tree using the parsimony optimality criterion.155 Thus, as phylogenetic principles underpin how we view genetic changes over time, One Health will also include the exchange of knowledge among evolutionary biologists and epidemiologists.

Phylogenetic trees are hard to compute and interpret. The need to consult professional phylogeneticists is made plain by the number of prominent papers that did not adhere to the standards of phylogenetics and failed to identify the fundamental hosts of coronaviruses.156 Moreover, a good phylogenetic analysis requires many elements: careful choice of the collected taxa, sequence, and or phenotypic data; method and quality control of sequence data and alignment; evaluation of substitution and indel models; treatment of partitions; tree-search protocol; measures of fit or confidence; and strategies for character coding and optimization.49,156,157 Moreover, results may vary with parameterization.158 These are only a few of the difficult decisions that go way beyond the level of sophistication of any software manuals and automated systems.156,159

**Are trees mapped to globes always needed?**

In many cases, such as the initial spread of H5N1 influenza, trees and Supramaps were very useful to understand the geographic spread of the pathogen, its multiple geographically and mutationally distinct patterns of zoonosis,18 and drug resistance.14 However, due to occlusion, Supramaps were not suitable for the visualization of cosmopolitan diseases, such as strains of *Salmonella* (eg, Hoffman et al3), seasonal influenza (eg, H3N2), pandemic influenza (H1N1-2009),16 and SARS-CoV-2. In response, researchers have worked on alternative visualization tools, including point-maps and route maps13,160 and eventually moved beyond the need for mapping trees to globes with Strainhub.161

Unlike Supramap, Strainhub is less computationally demanding. It can be executed from a web browser; it does not depend on closed source software (Google Earth), and geographical data are optional (Fig. 5). Moreover, Strainhub can be used to test hypotheses on the relative importance of hosts or places in disease spread. Future efforts for Strainhub will focus on usability, interoperability, visual clarity, and quantification of the relative importance of hosts or places in the spread of disease to better understand zoonosis.

**How do we prepare for the next pandemic?**

The COVID-19 pandemic has illustrated how unprepared our interconnected global society is for zoonotic disease. For the next pandemic, 2 frontiers of investigation are interesting for genomic epidemiology as a tool to survey microbes of pandemic potential to predict, prevent, or respond faster to the emergence of new disease.

First, we must survey the natural diversity of coronaviruses and other microbes of pandemic potential present within animals.120 Second, we must develop the science of pandemic prevention by moving from tracking pandemics that are occurring to predicting outbreaks. For example, combining artificial intelligence with genomic epidemiology can lead constructing a “viral forecast” to inform decisions about viruses with pandemic potential.162 Moreover, we have proposed a novel mathematical modeling framework based on agent-based modeling to predict pathogen patch dynamics underlying zoonosis.163

**Final remarks**

The COVID-19 pandemic, while ongoing, has caused 4,863,818 deaths worldwide as of October 14, 2021,164 and it has surpassed the US death toll from the 1918–1919 H1N1 pandemic, which was ∼675,000. As SARS-CoV-2 becomes endemic, we must remember that it is not as lethal as other pathogens such as H5N1 influenza or Nipah virus. In its last 100 years of existence, smallpox killed 300 million people, and *Variola major* (the major variant of smallpox) killed 30% of these patients.165

A novel pathogen at 30% mortality infecting 50% of the US population (166.7 million) would have resulted in 50 million deaths. MERS-CoV, henipaviruses, and hantavirus all have high mortality (>30%) and virulence with no approved vaccines or antivirals available. The 2018 Nipah outbreak had a 91% case-fatality rate, claiming 21 lives.146 We must heed the warning that pathogens with more severe disease phenotypes than SARS-CoV-2 could result in a far more devastating pandemic.
Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/ash.2021.222

Acknowledgments. We acknowledge the support of the following units of the University of North Carolina at Charlotte: the College of Computing and Informatics, the Bioinformatics Research Center, the Ribarsky Center for Visual Analytics, the Department of Bioinformatics and Genomics, Research and Economic Development, Academic Affairs, and University Research Computing. We further acknowledge the support of the North Carolina Research Campus and of the Belk Family. R.A. White III is supported by a UNC Charlotte start-up package. D.J.M. thanks Thiago José Jacob Carnevalli for his example and inspiration.

Financial support. No financial support was provided relevant to this article.

Conflict of interest. All authors report no conflicts of interest relevant to this article.

References

1. Janies DA. Phylogenetic concepts and tools applied to epidemiologic investigations of infectious diseases. Microbiol Spectr 2019;7: doi: 10.1128/microbiolspec.AME-0006-2018. Accessed November 10, 2021.
2. Janies DA, Pomeroy LW, Aaronson JM, et al. Analysis and visualization of H7 influenza using genomic, evolutionary and geographic information in a modular web service. Cladistics 2012;28:483–488.
3. Hoffmann M, Luo Y, Monday SR, et al. Tracing origins of the Salmonella Bareilly strain causing a foodborne outbreak in the United States. J Infect Dis 2016;213:502–508.
4. Ezeeoke I, Galac MR, Lin Y, et al. Tracking a serial killer: integrating phylogenetic relationships, epidemiology, and geography for two invasive meningooccal disease outbreaks. PLoS One 2018;13:e0202615.
5. Allard MW, Strain E, Melka D, et al. Practical value of food pathogen traceability through building a whole-genome sequencing network and database. J Clin Microbiol 2016;54: 1975–1983.
6. Marra MA, Jones SJM, Astell CR, et al. The genome sequence of the SARS-associated coronavirus. Science 2003;300:1399–1404.
7. Rota PA, Oberste MS, Monroe SS, et al. Characterization of a novel coronavirus associated with severe acute respiratory syndrome. Science 2003;300:1394–1399.
8. Boulos MNK. Descriptive review of geographic mapping of severe acute respiratory syndrome (SARS) on the Internet. Int J Health Geogr 2004;3:2.
9. Janies DA, Habib F, Alexandrov B, Hill A, Pol D. Evolution of genomes, host shifts and the geographic spread of SARS-CoV and related coronaviruses. Cladistics 2008;24:111–130.
10. Janies DA, Hill AW, Guralnick R, Habib F, Waltari E, Wheeler WC. Genomic analysis and geographic visualization of the spread of avian influenza (H5N1). Syst Biol 2007;56:321–329.
11. Janies DA, Treseder T, Alexandrov B, et al. The Supramap project: linking pathogen genomes with geography to fight emergent infectious diseases. Cladistics 2012;27:61–66.
12. Studer J, Janies DA. Global spread and evolution of viral haemorrhagic septicaemia virus. J Fish Dis 2011;34:741–747.
13. Janies DA, Voronkin IO, Studer J, et al. Selection for resistance to oseltamivir in seasonal and pandemic H1N1 influenza and widespread co-circulation of the lineages. Int J Health Geogr 2010;9:13.
14. Hill AW, Guralnick RP, Wilson MJ, Habib F, Janies D. Evolution of drug resistance in multiple distinct lineages of H5N1 avian influenza. Infect Genet Evol 2009;9:169–178.
15. Testimony of Daniel A. Janies, PhD. Local challenges of global proportions: evaluating role, preparedness for, and surveillance for pandemic onfluenza: Hearing before the committee on homeland security and government affairs, United States senate, 1 Sess. (2007). US government website. https://www.govinfo.gov/content/pkg/CHRG-110hrsg38846/html/CHRG-110hrsg38846.htm. Accessed October 7, 2021.
16. Janies DA, Voronkin IO, Das M, Hardman J, Treseder TW, Studer J. Genome informatics of influenza A: from data sharing to shared analytical capabilities. Anim Health Res Rev 2010;11:73–79.
17. Ruan YJ, Wei CL, Ee AL, et al. Comparative full-length genome sequence analysis of 14 SARS coronavirus isolates and common mutations associated with putative origins of infection. Lancet 2003;361:1779–1785.
18. Zhao G-P. SARS molecular epidemiology: a Chinese fairy tale of controlling an emerging zoonotic disease in the genomics era. Philos Trans R Soc Lond B Biol Sci 2007;362:1063–1081.
19. Reuter JA, Spacé D, Snyder MP. High-throughput sequencing technologies. Mol Cell 2015;58:586–597.
20. Pérez-Losada M, Arenas M, Galán JC, et al. High-throughput sequencing (HTS) for the analysis of viral populations. Infect Genet Evol 2020;80:104208.
21. The cost of sequencing a human genome. National Human Genome Research Institute website. https://www.genome.gov/about-genomics/fact-sheets-Sequencing-Human-Genome-cost. Accessed September 6, 2021.
22. Woo PCY, Lau SKP, Lam CSF, et al. Discovery of a novel bottlenose dolphin coronavirus reveals a distinct species of marine mammal coronavirus in gamma coronavirus. J Virol 2014;88:1318–1331.
23. Durães-Carvalho R, Caserta LC, Barnabé ACS, et al. Coronaviruses detected in Brazilian wild birds reveal close evolutionary relationships with beta- and deltacoronaviruses isolated from mammals. J Mol Evol 2015;81:21–23.
24. Woo PCY, Lau SKP, Lam CSF, et al. Discovery of seven novel mammalian and avian coronaviruses in the genus deltacoronavirus supports bat coronaviruses as the gene source of alphacoronavirus and betacoronavirus and avian coronaviruses as the gene source of gammacoronavirus and deltacoronavirus. J Virol 2012;86:3995–4008.
25. Woo PCY, Huang Y, Lau SKP, Yuen K-Y. Coronavirus genomics and bioinformatics analysis. Viruses. 2010;2:1804–1820.
26. Irigoyen N, Firth AE, Jones JD, Chung BY-W, Siddell SG, Briere I. High-resolution analysis of coronavirus gene expression by RNA sequencing and ribosome profiling. PLoS Pathog 2016;12:e1005473.
27. Coronavirusia. ViralZone website. https://viralzone.expasy.org/?sf=all_by_species. Accessed September 8, 2021.
28. Li BX, Ge JW, Li YJ. Porcine aminopeptidase N is a functional receptor for the PEDV coronavirus. Virology. 2007;365:166–172.
29. Boileau MJ, Kapil S. Bovine coronavirus associated syndromes. Vet Clin N Am Food Anim Pract 2010;26:123–146.
30. Mandelzik R, Sarvas M, Jackova A, Salamounova S, Novotny J, Vilcek S. First outbreak with chimeric swine enteric coronavirus (SeCoV) on pig farms in Slovakia—lessons to learn. Acta Vet Hung 2018;66:488–492.
31. Su S, Wong G, Shi W, Liu J, Lai ACK, Zhou J, et al. Epidemiology, genetic recombination, and pathogenesis of coronaviruses. Trends Microbiol 2016;24:490–502.
32. Lai MMC, Cavanagh D. The molecular biology of coronaviruses. In Advances in Virus Research. New York: Elsevier; 1997: 1–100.
33. Human coronavirus types. Centers for Disease Control and Prevention website. https://www.cdc.gov/coronavirus/types.html. Published March 17, 2021. Accessed September 7, 2021.
34. Zhang XM, Herbst W, Kousoulas KG, Storz J. Biological and genetic characterization of a hemagglutinating coronavirus isolated from a diarrhoeic child. J Med Virol 1994;44:152–161.
35. Zhong NS, Zheng BJ, Li YM, et al. Epidemiology and cause of severe acute respiratory syndrome (SARS) in Guangdong, People’s Republic of China, in February 2003. Lancet 2003;362:1353–1358.
36. Ksiazek TG, Erdman D, Goldsmith CS, et al. A novel coronavirus associated with severe acute respiratory syndrome. N Engl J Med 2003;348:1966–1970.
37. Zumla A, Hui DS, Perlman S. Middle East respiratory syndrome. Lancet 2015;386:995–1007.
38. World Health Organization director-general’s opening remarks at the media briefing on COVID–19— 11 March 2020. World Health Organization website. https://www.who.int/dg/speeches/detail/who-
director-general’s-opening-remarks-at-the-media-briefing-on-covid-19—11-march-2020. Accessed September 7, 2021.

39. Pekar J, Worobey M, Moshirt N, Scheffler K, Wertheim JO. Timing the SARS-CoV-2 index case in Hubei province. Science 2021;372:412–417.

40. Coronavirus (COVID-19) dashboard. World Health Organization website. https://covid19.who.int. Accessed September 18, 2021.

41. Yuen K-S, Ye Z-W, Fung S-Y, Chan C-P, Jin D-Y. SARS-CoV-2 and COVID-19: the most important research questions. Cell Biol 2020;10:40.

42. Andersen KG, Rambaut A, Lipkin WI, Holmes EC, Garry RF. The proximal origin of SARS-CoV-2. Nat Med 2020;26:450–452.

43. Liu S-L, Saif LJ, Weiss SR, Su L. No credible evidence supporting claims of the laboratory engineering of SARS-CoV-2. Emerg Microbes Infect 2020;9:505–507.

44. Holmes EC, Goldstein SA, Rasmussen AL, et al. The origins of SARS-CoV-2: a critical review. Cell 2021. doi: 10.1016/j.cell.2021.08.017.

45. Wang N, Li S-Y, Yang X-L, et al. Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis. J Travel Med 2020;27. doi: 10.1093/jtm/taaa022.

46. Shi Z-L. Origins of SARS-CoV-2: focusing on science. Infect Dis Immun 2021;1:3–4.

47. Machado DJ, Scott R, Guirales S, Janies DA. Fundamental evolution of all including three deadly lineages descendent from Chiroptera-hosted coronaviruses. SARS-CoV, MERS-CoV, and SARS-CoV-2. Cladistics 2021. doi: 10.1111/clad.12454.

48. Zhao S, Zhuang Z, Cao P, et al. Quantifying the association between domestic travel and the exportation of novel coronavirus (2019-nCoV) cases from Wuhan, China in 2020: a correlational analysis. J Travel Med 2020;27. doi: 10.1093/jtm/taaa022.

49. Li W, Shi Z, Yu M, et al. Bats are natural reservoirs of SARS-like coronaviruses. Science 2005;310:676–679.

50. Lytras S, Hughes J, Martin D, et al. Exploring the natural origins of SARS-CoV-2 in the light of recombination. bioRxiv 2021. doi: 10.1101/2021.01.22.427830.

51. Lytras S, Xia W, Hughes J, Jiang X, Robertson DL. The animal origin of SARS-CoV-2. Science 2021;373:968–970.

52. Lam TT-Y, Jia N, Zhang Y-W, et al. Identifying SARS-CoV-2-related coronaviruses in Malayan pangolins. Nature 2020;583:282–285.

53. Xiao K, Zhai J, Feng Y, et al. Isolation of SARS-CoV-2-related coronavirus from Malayan pangolins. Nature 2020;583:286–289.

54. Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr Biol 2020;30:1346–1351.e2.

55. Liu P, Jiang J-Z, Wan X-F, et al. Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? PLoS Pathog 2020;16:e1008421.

56. GISAID—Initiative. Global Initiative on Sharing All Influenza Data website. https://www.gisaid.org. Accessed September 18, 2021.

57. GISAID—Initiative. EpiCov website. https://www.epicov.org/. Accessed September 18, 2021.

58. Wu F, Zhao S, Yu B, et al. Author correction: a new coronavirus associated with human respiratory disease in China. Nature 2020;580:E7.

59. Sequencing of SARS-CoV-2. European Centre for Disease Control and Prevention website. https://www.ecdc.europa.eu/sites/default/files/documents/sequencing-of-SARS-CoV-2.pdf Accessed September 18, 2021.

60. Genomic sequencing of SARS-CoV-2: a guide to implementation for maximum impact on public health. World Health Organization website. https://www.who.int/publications/i/item/9789240018440. Published January 2021. Accessed September 18, 2021.

61. Truelove S, Smith CP, Qin M, et al. Projected resurgence of COVID-19 in the United States in July–December 2021 resulting from the increased transmissibility of the Delta variant and faltering vaccination. medRxiv 2021. doi: 10.1101/2021.08.28.21262748.

62. Lau BT, Pavlichin D, Hooker AC, et al. Profiling SARS-CoV-2 mutation fingerprints that range from the viral pangeneome to individual infection quasispecies. Genome Med 2021;13:62.

63. Hamza M, Ali A, Khan S, et al. nCoV-19 peptides mass fingerprinting identification, binding, and blocking of inhibitors flavonoids and anthraquinone of and hydroxychloroquine. J Biomed Struct Dyn 2021;39:4089–4099.

64. Hariton E, Locascio JJ. Randomised controlled trials—the gold standard for effectiveness research: Study design: randomised controlled trials. BJOG 2018;125:1716.

65. Boulware DR, Pullen MF, Bangdiwala AS, et al. A randomized trial of hydroxychloroquine as postexposure prophylaxis for COVID-19. N Engl J Med. 2020;383:517–525.

66. Siemieniuk RA, Bartoszek JJ, Ge L, et al. Drug treatments for COVID-19: living systematic review and network meta-analysis. BMJ 2020;370: m2980.

67. COVID-19 mRNA vaccine production. National Human Genome Research Institute website. https://www.genome.gov/about-genomics/fact-sheets/COVID-19-mRNA-Vaccine-Production. Accessed October 12, 2021.

68. Nakanishi N, Yoshio I. The novel coronavirus pandemic and the state of the epidemic in Kobe, Japan. J Disaster Res 2021;16:84–87.

69. Phillips N. The coronavirus is here to stay—here’s what that means. Nature 2021;590:382–394.

70. Zhou D, Dejnirattisai W, Supasa P, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine-induced sera. Cell 2021;184:2348–2361.e6.

71. Kustin T, Harel N, Finkel U, et al. Evidence for increased breakthrough rates of SARS-CoV-2 variants of concern in BNT162b2 mRNA-vaccinated individuals. Nat Med 2021;27:1379–1384.

72. Farinholt T, Doddapaneni H, Qin X, et al. Transmission event of SARS-CoV-2 delta variant reveals multiple vaccine breakthrough infections. BMC Med 2021;19:255.

73. Giovannetti M, Benedetti F, Campisi G, et al. Evolution patterns of SARS-CoV-2: snapshot on its genome variants. Biochem Biophys Res Comm 2021;538:88–91.

74. To KK-W, Sridhar S, Chiu KY-H, et al. Lessons learned 1 year after SARS-CoV-2 emergence leading to COVID-19 pandemic. Emerg Microbes Infect 2021;10:507–535.

75. Munafò MR, Tilling K, Taylor AE, Evans DM, Davey Smith G. Collider strata: When selection bias can substantially influence observed associations. Int J Epidemiol 2018;47:226–235.

76. Ramesh MA. Invited commentary: selection bias without colliders. Am J Epidemiol 2017;185:1048–1050.

77. Tattan-Birch H, Marsden J, West R, Gage SH. Assessing and addressing collider bias in addiction research: the curious case of smoking and COVID-19. Addiction 2021;116:982–984.

78. Brito AF, Semenova E, Dudas G, et al. Global disparities in SARS-CoV-2 genomic surveillance. medRxiv 2021. doi: 10.1101/2021.08.21.21262393.

79. Lauring AS, Hodcroft EB. Genetic variants of SARS-CoV-2: what do they mean? JAMA 2021;325:529–531.

80. Legally H, Wilkinson E, Giovannetti M, 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 2020. doi: 10.1101/2020.12.21.20248640.

81. Ford CT, Scott R, Machado DJ, Janies D. Sequencing data of North American SARS-CoV-2 isolates shows widespread complex variants. medRxiv 2021. doi: 10.1101/2021.01.27.21250648.

82. Awadassleid A, Wu Y, Tanaka Y, Zhang W. Current advances in the development of SARS-CoV-2 vaccines. Int J Biol Sci 2021;17:8–19.

83. Hossain MK, Hassanadeganroudsi M, Apostolopoulos V. The emergence of new strains of SARS-CoV-2. What does it mean for COVID-19 vaccines? Expert Rev Vaccines 2021;20:635–638.

84. Ul-Rahman A, Shabbir MAB, Aziz MW, et al. A comparative phylogenomic analysis of SARS-CoV-2 strains reported from non-human
mammalian species and environmental samples. Mol Biol Rep 2020;47:9207–9217.
89. Kuhn JH, Bao Y, Bavari S, et al. Virus nomenclature below the species level: a standardized nomenclature for natural variants of viruses assigned to the family Filoviridae. Arch Virol 2013;158:301–311.
90. Parumus D. Editorial: revised World Health Organization (WHO) terminology for variants of concern and variants of interest of SARS-CoV-2. Med Sci Monit 2021;27:e933622.
91. Konings F, Perkins MD, Kuhn JH, et al. SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. Nat Microbiol 2021;6:821–823.
92. Janik E, Niemczewicz M, Podogrocki M, Majsterek I, Bijak M. The emerging concern and interest SARS-CoV-2 variants. Pathogens 2021;10. doi: 10.3390/pathogens10060633.
93. Tracking SARS-CoV-2 variants. World Health Organization website. https://www.who.int/activities/tracking-SARS-CoV-2-variants. Accessed September 20, 2021.
94. SARS-CoV-2 variant classifications and definitions. Centers for Disease Control and Prevention website. https://www.cdc.gov/coronavirus/2019-ncov/variants/variant-info.html. Published 2021. Accessed September 20, 2021.
95. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data from vision to reality. Euro Surveill 2017;22. doi: 10.2807/1560-7917.ES.2017.22.13.304943.
96. Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018;34:4121–4123.
97. Rabmuth A, Holmes EC, O’Toole À, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol 2020;5:1403–1407.
98. Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol 2021;19:409–424.
99. Flanagan KL, Macintyre CR, McIntyre PB, Nelson MR. SARS-CoV-2 vaccine: where are we now? J Allergy Clin Immunol Pract 2021. doi: 10.1016/j.jaip.2021.07.016.
100. Cevik M, Grubauh ND, Iwasaki A, Openshaw P. COVID-19 vaccines: keeping pace with SARS-CoV-2 variants. Cell 2021. doi: 10.1016/j.cell.2021.09.010.
101. Schmitz AJ, Turner JS, Liu Z, et al. A vaccine-induced public antibody protects against SARS-CoV-2 and emerging variants. Immunology 2021;54:2159–2166.
102. Boehm E, Kronig I, Neher RA, et al. Novel SARS-CoV-2 variants: the pandemics within the pandemic. Clin Microbiol Infect 2021;27:1109–1117.
103. Farooqi T, Malik JA, Mullah AH, et al. An overview of SARS-CO-2 epidemiology, mutant variants, vaccines, and management strategies. J Infect Public Health 2021. doi: 10.1016/j.jiph.2021.08.014.
104. Krause PR, Gruber MF. Emergency use authorization of COVID vaccines —safety and efficacy follow-up considerations. N Engl J Med 2020;383: e107.
105. Pascual-Iglesias A, Canton J, Ortega-Prieto AM, Jimenez-Guardaño JM, Regla-Nava JA. An overview of vaccines against SARS-Cov-2 in the COVID-19 pandemic era. Pathogens 2021;10. doi: 10.3390/pathogens10081030.
106. Chen Y, Zhu L, Huang W, et al. Potent RBD-specific neutralizing rabbit monoclonal antibodies recognize emerging SARS-CoV-2 variants elicited by DNA prime-protein boost vaccination. Emerg Microbes Infect 2021;10: 1390–1403.
107. Hodgson SH, Mansatta K, Mallett G, Harris V, Emary KRW, Pollard AJ. What defines an efficacious COVID-19 vaccine? A review of the challenges assessing the clinical efficacy of vaccines against SARS-CoV-2. Lancet Infect Dis 2021;21:e26–e35.
108. Zhao T, Hu C, Ayaz Ahmed M, Cheng C, Chen Y, Sun C. Warnings regarding the potential coronavirus disease 2019 (COVID-19) transmission risk: vaccination is not enough. Infect Control Hosp Epidemiol 2021;2: 377–390.
134. Kim Y-I, Kim S-G, Kim S-M, et al. Infection and rapid transmission of SARS-CoV-2 in ferrets. Cell Host Microbe 2020;27:704–709.e2.

135. Freuling CM, Breithaupt A, Müller T, et al. Susceptibility of raccoon dogs for experimental SARS-CoV-2 infection. Emerg Infect Dis 2020;26:2982–2985.

136. Munster VJ, Feldmann F, Williamson BN, et al. Respiratory disease in rhesus macaques inoculated with SARS-CoV-2. Nature 2020;585:268–272.

137. Rocks B, Kuiken T, Herfst S, et al. Comparative pathogenesis of COVID-19, MERS, and SARS in a nonhuman primate model. Science 2020;368:1012–1015.

138. Mykytyn AZ, Lamers MM, Okba NMA, et al. Susceptibility of rabbits to SARS-CoV-2. Emerg Microbes Infect 2021;10:1–7.

139. Schlottau K, Rissmann M, Graaf A, et al. SARS-CoV-2 in fruit bats, ferrets, pigs, and chickens: an experimental transmission study. Lancet Microbe 2020;1:e218–e225.

140. Imai M, Iwatsuki-Horimoto K, Hatta M, et al. Syrian hamsters as a small animal model for SARS-CoV-2 infection and countermeasure development. Proc Nat Acad Sci 2020;117:16387–16395.

141. Palmer MV, Martins M, Falkenberg S, et al. Susceptibility of white-tailed deer (Odocoileus virginianus) to SARS-CoV-2. J Virol 2021. doi: 10.1128/JVI.00083-21.

142. Chandler JC, Bevins SN, Ellis JW, et al. SARS-CoV-2 exposure in wild white-tailed deer (Odocoileus virginianus). bioRxiv 2021. doi: 10.1101/2021.07.29.454326.

143. Gryseels S, De Bruyn L, Gyselings R, Calvignac-Spencer S, Leendertz FH, Leirs H. Risk of human-to-wildlife transmission of SARS-CoV-2. Mamm Rev 2020. doi: 10.1111/mam.12225.

144. Griffin JB, Haddix M, Danza P, et al. SARS-CoV-2 Infections and hospitalizations among persons aged ≥16 years, by vaccination status—Los Angeles County, California, May 1–July 25, 2021. Morb Mortal Wkly Rep 2021;70:1170–1176.

145. Del Rio C, Malani PN, Omer SB. Confronting the delta variant of SARS-CoV-2, summer 2021. JAMA 2021. doi: 10.1001/jama.2021.14811.

146. Lazarevic I, Pravica V, Miljanovic D, Cupic M. Immune evasion of SARS-CoV-2 emerging variants: what have we learnt so far? Viruses 2021;13. doi: 10.3390/v13071192.

147. Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature 2021;592:277–282.

148. Srivastava S, Banu S, Singh P, Sowpati DT, Mishra RK. SARS-CoV-2 genomics: an Indian perspective on sequencing viral variants. J Biosci 2021;46. doi: 10.1007/s12038-021-00145-7.

149. Furuse Y. Genomic sequencing effort for SARS-CoV-2 by country during the pandemic. Int J Infect Dis 2021;103:305–307.

150. Crawford DC, Williams SM. Global variation in sequencing impedes SARS-CoV-2 surveillance. PLoS Genet 2021;17:e1009620.

151. Adepoju P. Challenges of SARS-CoV-2 genomic surveillance in Africa. Lancet Microbe 2021;2:e139.

152. Blomberg N, Lauer KB. Connecting data, tools and people across Europe: ELIXIR’s response to the COVID-19 pandemic. Eur J Hum Genet 2020;28:719–723.

153. Conesa A, Beck S. Making multimics data accessible to researchers. Sci Data 2019;6:251.

154. Hodcroft EB, De Maio N, Lanfear R, et al. Want to track pandemic variants faster? Fix the bioinformatics bottleneck. Nature 2021;591:30–33.

155. Turakhia Y, Thornlow B, Hinrichs AS, et al. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. Nat Genet 2021;53:809–816.

156. Wenzel J. Origins of SARS-CoV-1 and SARS-CoV-2 are often poorly annotated for viral genomes of Flaviviridae. Viruses 2020;12. doi: 10.3390/v12080892.

157. Wheeler WC. Sequence alignment, parameter sensitivity, and the phylogenetic analysis of molecular data. Syst Biol 1995;44:321–331.

158. Grant T. The perils of “point-and-click” systems. Cladistics 2003;19:276–285.

159. Hovmöller R, Alexandrov B, Hardman J, Janies D. Tracking the geographical spread of avian influenza (H5N1) with multiple phylogenetic trees. Cladistics 2010;26:1–13.

160. de Bernardi Schneider A, Ford CT, et al. StrainHub: a phylogenetic tool to construct pathogen transmission networks. Bioinformatics 2020;36:945–947.

161. Syrowatka A, Kuznetsova M, Alsubai A, et al. Leveraging artificial intelligence for pandemic preparedness and response: a scoping review to identify key use cases. NPJ Digit Med 2021;4:96.

162. Chen S, Owolabi Y, Li A, et al. Patch dynamics modeling framework from pathogens’ perspective: Unified and standardized approach for complicated epidemic systems. PLoS One 2020;15:e0238186.

163. COVID-19. Johns Hopkins Coronavirus Resource Center website. https://coronavirus.jhu.edu/map.html. Accessed October 13, 2021.

164. Henderson DA. The eradication of smallpox—an overview of the past, present, and future. Vaccine 2011;29 suppl 4:S7–9.

165. Arunkumar G, Chandni R, Mourya DT, et al. Outbreak investigation of Nipah virus disease in Kerala, India, 2018. J Infect Dis 2019;219:1867–1878.

166. Soh SM, Kim Y, Kim C, Jang US, Lee H-R. The rapid adaptation of SARS-CoV-2-α of the variants: transmission and resistance. J Microbiol 2021;59:807–818.

167. Schwarze K, Buchanan J, Fermont JM, et al. The complete costs of genome sequencing: a microcosting study in cancer and rare diseases from a single center in the United Kingdom. Genet Med 2020;22:85–94.