CellPress
REVIEWS

[1]Department of Organismic and Evolutionary Biology, Harvard University, Cambridge, MA 02138, USA

*Correspondence:
danielpark@fas.harvard.edu (D.S. Park).

References

1. van Dorn, A. *et al.* (2020) COVID-19 exacerbating inequalities in the US. *Lancet* 395, 1243–1244
2. Corbera, E. *et al.* (2020) Academia in the time of COVID-19: towards an ethics of care. *Plan. Theory Pract.* Published online May 18, 2020. https://doi.org/10.1080/14649357.2020.1757891
3. Curtis, R.B. and Council, N.R. (1969) *The Invisible University: Postdoctoral Education in the United States: Report of a Study Conducted under the Auspices of the National Research Council*, National Academy of Sciences
4. Sauermann, H. and Roach, M. (2016) Why pursue the postdoc path? *Science* 352, 663–664
5. Yan, W. (2020) Early-career scientists at critical career junctures brace for impact of COVID-19. *Science.* Published online April 7, 2020. https://doi.org/10.1126/science.caredit.abc1291
6. Shen, H. (2015) Employee benefits: plight of the postdoc. *Nature* 525, 279–281
7. Ålund, M. *et al.* (2020) Academic ecosystems must evolve to support a sustainable postdoc workforce. *Nat. Ecol. Evol.* 4, 777–781
8. Ledford, H. (2017) US postdocs face steep challenges when starting families. *Nat. News.* Published online June 29, 2017. https://www.nature.com/news/us-postdocs-face-steep-challenges-when-starting-families-1.22200
9. National Academy of Sciences (2014) *The Postdoctoral Experience Revisited*, National Academies Press
10. Servick, K. *et al.* (2020) Coronavirus disruptions reverberate through research. *Science* 367, 1289–1290
11. Middleton, C. (2020) Pandemic creates hurdles for academic job seekers. *Physics Today.* Published online April 23, 2020. https://doi.org/10.1063/PT.6.2.20200423a
12. Woolston, C. (2019) Why a postdoc might not advance your career. *Nature* 565, 125–127
13. Stephan, P. (2013) How to exploit postdocs. *Bioscience* 63, 245–246
14. National Research Council (US) Committee on Bridges to Independence: Identifying Opportunities for and Challenges to Fostering the Independence of Young Investigators in the Life Sciences (2005) *Bridges to Independence: Fostering the Independence of New Investigators in Biomedical Research*, National Academies Press
15. Brooks, I.M. (2008) Postdoc glut means academic pathway needs an overhaul. *Nature* 454, 1049

Spotlight

Tracking the Genomic Footprints of SARS-CoV-2 Transmission

Tommy Tsan-Yuk Lam[1],*

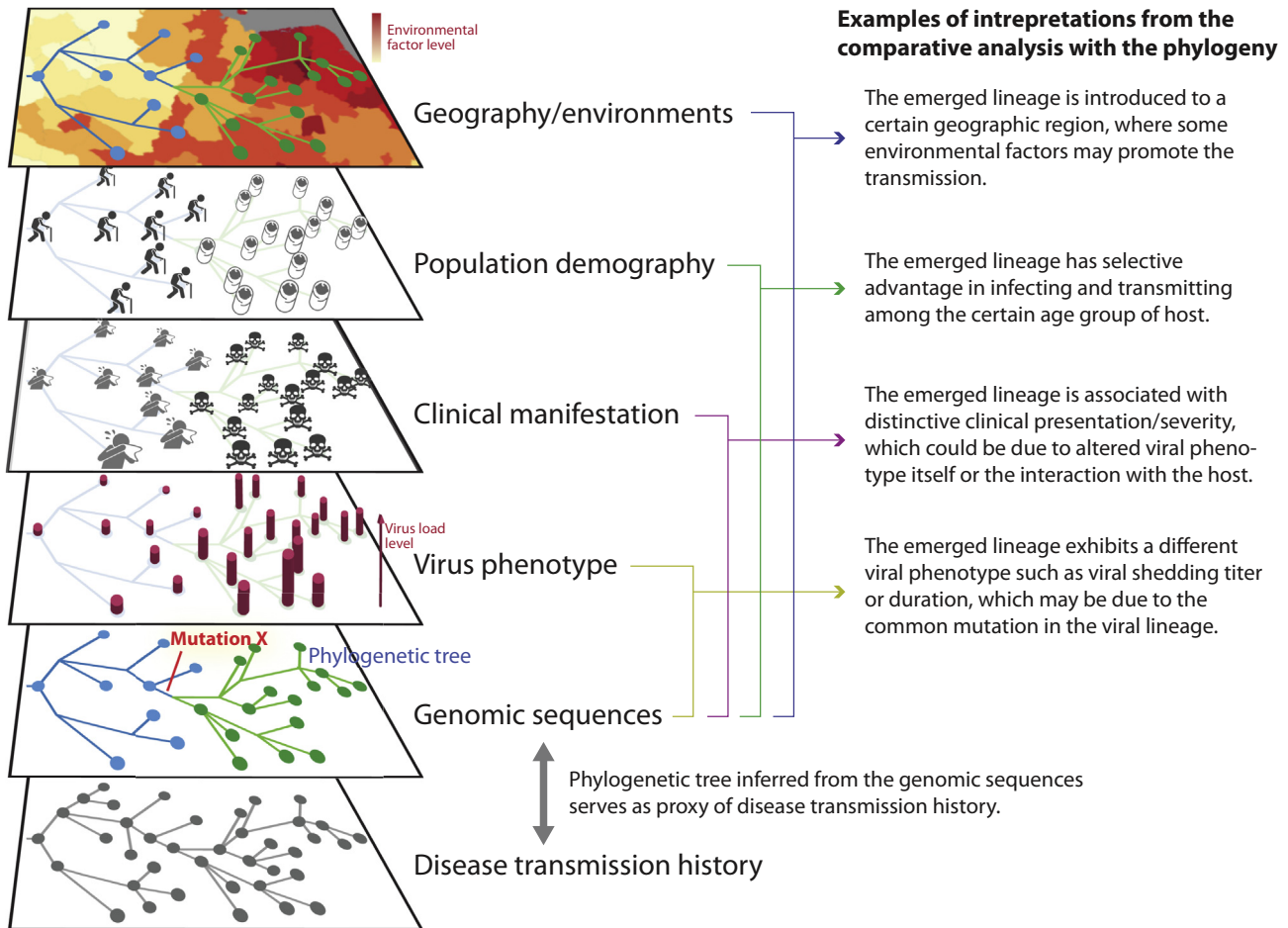**There is considerable public and scientific interest in the origin, spread, and evolution of SARS-CoV-2. Lu *et al.* recently conducted genomic sequencing and analysis of SARS-CoV-2 in Guangdong, revealing its early transmission out of Hubei and shedding light on the effectiveness of controlling local transmission chains.**

The outbreak of SARS-CoV-2 (the cause of the disease known as COVID-19) was first reported in Wuhan city (Hubei province), China, in December 2019, and swept across the nation and then around the world within 2 months, leading to its declaration as a pandemic by the World Health Organisation (WHO) on 11 March 2020. As of 20 May 2020, the worldwide COVID-19 death toll has reached 318 789, and nearly 5 million confirmed cases have been reported across all continents [1]. The enormous health and economic impacts caused by this virus, at such an unprecedented speed, have raised considerable public interest, fueling research into its origin, spread, and evolution. Analysis of genomic data is a key tool for studying emerging pathogens [2]. The recent identification and genomic characterization of bat and pangolin coronaviruses, which are evolutionarily related to human SARS-CoV-2, have suggested that these species could act as the zoonotic origins of the human virus [3,4]. Other pressing questions for understanding the ongoing pandemic include identifying the means by which SARS-CoV-2 has spread across China and the world from its starting place, and whether disease control measures effectively suppress the introduced infections against further transmission.

Several studies using mathematical modeling of COVID-19 and other coronavirus transmission have suggested that measures such as social distancing and city lockdown may be effective methods of disease control [5,6], but these seldom assess individual transmission chains. Lu *et al.* recently sequenced SARS-CoV-2 genomes from 53 patients in Guangdong, the province of China that reported the highest number of cases (*n* = 1388) outside Hubei [7]. Phylogenetic analysis of these and other genomes of SARS-CoV-2 worldwide revealed that the Guangdong virus strains are scattered among strains from other Chinese provinces and other countries, consistent with the epidemiological finding that the majority of infections detected in Guangdong were imported cases. Twenty-three Guangdong virus strains formed five statistically well-supported phylogenetic clusters, and this might indicate some local transmission following the introductions of these virus lineages. Notably, the observed durations of these putative local transmission chains largely overlap with the period from mid-January to late February in which most Guangdong local infections defined by the patients' travel histories were reported. These genetic and epidemiologic findings consistently suggest that control measures successfully reduced local transmission in Guangdong after February.

Genomic information about pathogens provides valuable empirical information about their transmission histories [1], such as the identification of transmission chains through phylogenetic analysis of genome sequences, as illustrated by the work of Lu *et al.* [7]. In addition to revealing pathogen evolutionary processes and acting as a proxy of disease transmission history, phylogenetic trees also serve as versatile frameworks for comparative analysis of virus genetics and phenotypes, disease epidemiology, clinical manifestations, and population demography and environments, thus facilitating the identification of possible interplay between these various aspects of disease dynamics (Figure 1). There has been active research into methods for integrating multidimensional data related to pathogens for statistical inference, especially in a Bayesian framework [8]. It is anticipated that such integrative analysis of SARS-CoV-2, including the work by Lu *et al.* [7], as well as a more recent study by Dellicour *et al.* in Belgium [9], will greatly promote

Examples of intrepretations from the comparative analysis with the phylogeny

Geography/environments — The emerged lineage is introduced to a certain geographic region, where some environmental factors may promote the transmission.

Population demography — The emerged lineage has selective advantage in infecting and transmitting among the certain age group of host.

Clinical manifestation — The emerged lineage is associated with distinctive clinical presentation/severity, which could be due to altered viral phenotype itself or the interaction with the host.

Virus phenotype — The emerged lineage exhibits a different viral phenotype such as viral shedding titer or duration, which may be due to the common mutation in the viral lineage.

Genomic sequences

Phylogenetic tree inferred from the genomic sequences serves as proxy of disease transmission history.

Disease transmission history

Trends in Genetics

Figure 1. Illustration of Comparative Analysis of Different Types of Data Related to Emerging Infections. The disease transmission chain in an outbreak is depicted in the last layer. The penultimate layer shows the genomic samples of the pathogen from the outbreak which are connected in a phylogenetic tree. A newly emerged lineage is highlighted in green and is characterized by the mutation X in red. Other related data such as virus phenotypes, clinical manifestations, population demography, geography, and environmental features can be mapped to the phylogenetic tree based on the information associated with the genomic samples, as shown in the first to fourth (from top down) layers. Comparison to the genomic sequences in the phylogenetic tree can identify the association with the emerged lineage (green), and some examples of such an interpretation are indicated in the text in the right column. Notably, an association may also be established between two or more layers – for example, the emerged lineage carries mutation X (fifth layer), which increases virus shedding titer and duration of infection (fourth layer), and hence causes higher disease severity (third layer).

our understanding of virus dynamics and interactions with hosts and environments, and pave the way to genomic surveillance studies of SARS-CoV-2 in the near future. Despite the usefulness of genomic data and phylogenetic analysis, the resolution of the SARS-CoV-2 phylogeny in the early phase of the pandemic remains low, possibly because of the relatively slow genetic drift of the virus, as well as differential sampling intensities in different regions. Many lineages are defined by a single mutation, and hence the phylogenetic structure could easily be distorted by sequences that carry a high number of sequencing errors or mutations introduced during passage in virus culture. One important issue is the variable intensity of virus genome sampling in different regions. This is seen in the SARS-CoV-2 dataset used by Lu et al. [7], which has relatively few genomes from outside Guangdong province (e.g., only 32 genomes from Hubei province, which had >60000 cases at the time). Such undersampling in regions with high disease incidence could result in phylogenetic analyses that underestimate virus exportation from these regions to other regions [10] such as Guangdong, and hence the extent of local transmission in Guangdong might be overestimated [7]. Notably, the SARS-CoV-2 genome dataset publicly available at time of writing remains highly uneven across different countries. For instance, the GISAID

(Global Initiative on Sharing All Influenza Data; https://www.gisaid.org) database had only five full genome sequences from Iran where >120000 cases have been reported; by contrast, it has ~1300 genomes from ~7000 cases reported in Australia. Therefore, any interpretation of phylogenetic analyses involving undersampled regions must be made with caution. These limitations constrain certainty in interpretations obtained using phylogenetics, and a conservative approach is essential when interpreting results from complex phylogenetic models and multidimensional data.

Many countries are now investing efforts into genomic surveillance of SARS-CoV-2, and data sharing on the GISAID public database has now reached 25995 full genomes at unprecedented speed. Although the challenges of low phylogenetic resolution and biased sampling will undoubtedly remain, it is anticipated that future research into these SARS-CoV-2 genomes will take on these challenges with more robust statistical methods and cautious data interpretation, and will potentially provide important insights into SARS-CoV-2 transmission and evolution within different countries and across the world thereby aiding more effective control of the disease.

[1]State Key Laboratory of Emerging Infectious Diseases, School of Public Health, The University of Hong Kong, Hong Kong SAR, China

*Correspondence:
ttylam@hku.hk (T.T.-Y. Lam).

References
1. World Health Organization (2020) *Coronavirus Disease (COVID-2019) Situation Report 121*, WHO
2. Pybus, O.G. and Rambaut, A. (2009) Evolutionary analysis of the dynamics of viral infectious disease. *Nat. Rev. Genet.* 10, 540–550
3. Zhou, P. *et al.* (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* 579, 270–273
4. Lam, T.T. *et al.* (2020) Identifying SARS-CoV-2 related coronaviruses in Malayan pangolins. *Nature* Published online March 26, 2020. https://doi.org/10.1038/s41586-020-2169-0
5. Kraemer, M.U.G. *et al.* (2020) The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* 368, 493–497
6. Kissler, S.M. *et al.* (2020) Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* 368, 860–868
7. Lu, J. *et al.* (2020) Genomic epidemiology of SARS-CoV-2 in Guangdong province, China. *Cell* 181, 997–1003.e9
8. Suchard, M.A. *et al.* (2018) Bayesian phylogenetic and phylodynamic data integration using BEAST 1.10. *Virus Evol.* 4, vey016
9. Dellicour, S. *et al.* (2020) A phylodynamic workflow to rapidly gain insights into the dispersal history and dynamics of SARS-CoV-2 lineages. *BioRxiv* Published online May 9, 2020. https://doi.org/10.1101/2020.05.05.078758
10. Grubaugh, N.D. *et al.* (2017) Genomic epidemiology reveals multiple introductions of Zika virus into the United States. *Nature* 546, 401–405

Spotlight

SpRY: Engineered CRISPR/Cas9 Harnesses New Genome-Editing Power

Dangquan Zhang[1],* and Baohong Zhang[2],*

Due to protospacer adjacent motif (PAM) requirements, CRISPR/Cas9 cannot access many genetic loci. A recent study by Walton *et al.* structurally engineered *Streptococcus pyogenes* Cas9 (SpCas9) to near-PAMless SpRY that can target most DNA sequences with high editing efficiency and flexibility. This newly engineered SpRY will potentially expand genome-editing capabilities for basic and applied research.

PAM Requirements Limit the Potentials of Genome Editing

CRISPR/Cas9-based genome editing affords scientists the ability to precisely edit certain genes in a genome. Since its usability was first recognized, the CRISPR/Cas9 system has quickly become a genome-editing tool, widely used to edit specific DNA sequences across nearly all organisms [1]. However, traditional CRISPR/Cas9-based genome editing requires a PAM in the edited gene that serves as the recognition site for Cas9 enzymes. Although different Cas variants recognize different PAM sequences, the most commonly used SpCas9 recognizes NGG sequences; CRISPR/Cas9 has been widely used to edit and monitor a specific DNA sequence containing the NGG PAM sequence for a variety of purposes [2]. However, the required PAM sequences limit the sequences that CRISPR/Cas9 can target. Scientists have successfully identified Cas enzymes with differing PAMs that expand the Cas-targeted DNA sequences; however, further strategies are warranted for wider application of genome editing in gene function studies and beyond.

To relax the PAM requirements, scientists have modified Cas enzymes to recognize a wider range of PAMs using several strategies, including directed evolution and structure-guided engineering. Modified Cas9s, such as SpCas9-NG and SpCas9-VQR, have relaxed PAM preferences and recognize altered PAM sequences [3–5]. Although these modifications have expanded DNA target sites, these sites are still very limited, and most DNA sequences remain inaccessible for CRISPR/Cas-based genome editing.

Engineered CRISPR/Cas9s Relax the PAM Requirements

A recent study by Walton and colleagues [6] may completely change this field, making CRISPR/Cas9 available to edit genes at any DNA sequence. In their study, Walton and colleagues genetically modified the SpCas9 enzyme through a serial structure-guided engineering approach. They first generated a modified SpCas9 variant, named SpG, to alter the SpCas9 PAM preference and relax the Cas9 recognition sites; the newly modified SpG could target an expanded set of NGN PAMs instead of only NGG PAM. This change made CRISPR/Cas9 capable of editing four times the number of genes over that of the original system (Figure 1). Compared with the traditional