Five Challenges in the Field of Viral Diversity and Evolution

Rafael Sanjuán[1]*, Christopher J. R. Illingworth[2], Jemma L. Geoghegan[3,4], Jaime Iranzo[5,6], Mark P. Zwart[7], Alexander T. Ciota[8], Gonzalo Moratorio[9,10], Selma Gago-Zachert[11], Siobain Duffy[12] and Dhanasekaran Vijaykrishna[13]

[1] Institute for Integrative Systems Biology (I2SysBio), Consejo Superior de Investigaciones Científicas (CSIC), Universitat de València, València, Spain, [2] Medical Research Council Biostatistics Unit, University of Cambridge, Cambridge, United Kingdom, [3] Department of Microbiology and Immunology, University of Otago, Dunedin, New Zealand, [4] Institute of Environmental Science and Research, Wellington, New Zealand, [5] Centro de Biotecnología y Genómica de Plantas, Universidad Politécnica de Madrid (UPM), Madrid, Spain, [6] Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria (INIA), Madrid, Spain, [7] Department of Microbial Ecology, Netherlands Institute of Ecology, The Royal Netherlands Academy of Arts and Sciences, Wageningen, Netherlands, [8] Wadsworth Center, New York State Department of Health, Albany, NY, United States, [9] Laboratorio de Evolución Experimental de Virus, Institut Pasteur de Montevideo, Montevideo, Uruguay, [10] Laboratorio de Virología Molecular, Facultad de Ciencias, Universidad de la República, Montevideo, Uruguay, [11] Section Microbial Biotechnology, Institute of Biochemistry and Biotechnology, Martin Luther University Halle-Wittenberg, Halle, Germany, [12] Department of Ecology, Evolution, and Natural Resources, Rutgers University, New Brunswick, NJ, United States, [13] Division of Public Health Laboratory Sciences, Hong Kong University, Hong Kong, China

SUMMARY

Viral diversity and evolution play a central role in processes such as disease emergence, vaccine failure, drug resistance, and virulence. However, significant challenges remain to better understand and manage these processes. Here, we discuss five of these challenges. These include improving our ability to predict viral evolution, developing more relevant experimental evolutionary systems, integrating viral dynamics and evolution at different scales, systematic appraisal of the virosphere, and deepening our understanding of virus-virus interactions. Intensifying future research on these areas should improve our ability to combat viral diseases, as well as to exploit viral diversity and evolution for biotechnological purposes.

INTRODUCTION

Viruses are extremely abundant, genetically diverse, and fast-evolving entities. They possess widely different replication strategies, genome organizations, and virion structures. Many viruses, particularly RNA viruses, evolve within short, human-observable timescales (1). This helps us not only understand previous evolutionary events, but potentially also predict their occurrence, anticipating important processes such as virulence evolution, vaccine escape, and drug resistance. For instance, the recent SARS-CoV-2 pandemic highlights the need to understand patterns of viral cross-species transmission, emergence, diversification, and adaptation to a novel host. However, evolution is an open-ended, complex, and inherently stochastic process, challenging our ability to predict its course. Multiple challenges therefore remain. For example, it is often found that the conclusions obtained about viral evolution in one context or timescale are no longer correct in alternative settings, for instance when short-term evolutionary rates are extrapolated to longer timescales. Achieving greater generality and consistency requires the collection of more comprehensive data and the construction of more flexible and integrative models for its interpretation. Below, we discuss specific challenges underpinning the goal of achieving higher predictability, consistency, and comprehensiveness in the field.

ANTICIPATING VIRAL EVOLUTION

Predicting viral evolution is a long-sought goal, a particular focus of such efforts being human strains of influenza virus, which require constant vaccine updates (2). The advent of high-throughput sequencing has facilitated massive genome-wide viral sequencing in real time during the course of viral epidemics, as already evidenced during the 2013–2014 Western African Ebola epidemic (3) and the emergence of Zika in South America in 2015–2016 (4). With the SARS-CoV-2 pandemics in 2020, this approach reached massive proportions, with hundreds of thousands of full-genome sequences being released within months (gisaid.org) (5). By linking epidemiological information with sequence data, it has been possible to track the early spread of new variants exhibiting enhanced transmissibility, such as lineage B.1.1.7 emerged in the UK, anticipating their ability to outcompete other lineages and become dominant in different populations (6). Data collection on this scale constitutes a turning point in the field, making it feasible to predict certain aspects of viral evolution in the real world at a global scale.

However, major challenges still remain in this area. Although we are currently capable of forecasting the dynamics of some viral variants, our ability to predict the emergence of new variants of concern is still somewhat limited. Progress in this direction will require additional tools, including the experimental evaluation of fitness landscapes and mutational pathways to assess the evolutionary potential of a viral strain (7). RNA viruses are particularly convenient systems for investigating this topic, since they have simple genomes, can be manipulated easily, and mutate rapidly. For instance, the recurrent emergence of certain variants both in laboratory evolution experiments and in nature can help us predict the evolution of some viral features (8, 9).

Nevertheless, we are still far from reaching predictability in other settings, particularly concerning viral transmission from wild animals into humans, which is the main source of new, potentially pandemic viruses. With hundreds of thousands of wildlife viruses with zoonotic potential but just a handful of zoonotic events, it is at present not possible to say when and where a new virus will first infect and spread efficiently among humans (10). However, risk factors such as local biodiversity, ecological disturbance, viral genetic features, receptor usage, and host tropism have been identified (11, 12). In the not-so-distant future, we might achieve better predictability by combining viromics, machine learning, and experimental virology tools (13–15).

DEVELOPING MORE RELEVANT EXPERIMENTAL EVOLUTION SYSTEMS

Controlled laboratory experiments can provide useful information about the repeatability and predictability of viral evolution (16). Surprisingly, non-viral systems have often predominated in this discussion (17), despite the suitability of viruses, considering their fast evolution and simple genomes. It has been well-established that, under strong directional selection, viral evolution in the laboratory often produces parallel evolution events at the sequence level, as for instance shown in a recent long-term HIV evolution experiment (18). This information can be used for identifying candidate adaptive mutations, which can be then analyzed by genetic engineering. Advances such as deep mutational scanning and microfluidics help us generate viral diversity more efficiently in the laboratory, and allow experimental evolution to be carried out in a more systematic and automated manner, accelerating adaptation and increasing experimental repeatability (7).

By combining laboratory virus evolution with high-throughput mutagenesis or sequencing, virologists have characterized the distribution of mutational fitness effects with unprecedented detail, such that the effect of nearly every possible individual mutation can now be measured (19). However, the question remains to what extent these experiments, which are typically carried out in simple cell culture systems, capture important aspects of viral evolution in nature. Certain selective pressures, such as receptor usage, are probably well-modeled using cell cultures. However, evolutionary experiments are typically conducted using monolayers of tumoral cells, which tend to exhibit aberrant innate immune responses and thus offer an unnaturally permissive environment for viral spread. Cultures of non-tumoral cells, as well as 3D cultures, organoids, and explants should better reflect some selective pressures found *in vivo* (20, 21). On the other hand, these systems are harder to implement and less efficient at selecting specific viral mutations because they achieve lower viral yields and, consequently, lower effective population sizes. One way or another, we need to better link experimental evolution systems with virus diversity and evolution in nature by contrasting them with field data. Finally, we should acknowledge that, at present, most viruses identified in nature have not even been isolated, and thus cannot be cultured. Despite these major limitations, experimental strategies such as creating pseudotypes or replicons by *de novo* DNA or RNA synthesis help us investigate important aspects of some wildlife viruses in the laboratory, including receptor-dependent cell tropism and innate immunity evasion.

INTEGRATING VIRAL DYNAMICS AND EVOLUTION AT DIFFERENT SCALES

An important issue in the field is that viral evolution rates inferred from phylogenetic analyses are strongly dependent on measurement timescales, such that rates estimated from recent isolates (e.g., seasonal outbreaks) are systematically higher than those estimated for longer time periods [e.g., permafrost or archival samples; (22, 23)]. This inconsistency needs to be addressed for viral phylogenetic trees to provide more reliable information about the time at which different groups evolved. Similarly, to date relatively few studies have successfully linked evolutionary processes occurring at different levels, spanning intra-host, inter-host, and community levels. Some important steps forward have been recently made in this direction, though, such as identifying common evolutionary events at these different spatiotemporal scales (24).

A particularly challenging issue is defining and measuring viral fitness at these different organizational levels. Whereas at certain stages of infection or in a simplified laboratory setting fitness is largely determined by the rate at which a virus can infect cells (which in turn includes factors such as receptor binding, replication and gene expression rates, etc.), the ability to evade innate and adaptive immunity becomes increasingly important as the infection progresses. In the process of transmission, other factors such as the ability to be shed from hosts and environmental stability become critical. There is a need to better link mechanistically well-defined aspects of the viral infection cycle (e.g., polymerase catalytic efficiency, receptor binding affinity, capsid stability, etc.) to observable fitness at the population level (e.g., the basic reproductive rate, R_0).

CHARACTERIZING AND UNDERSTANDING THE VIROSPHERE MORE THOROUGHLY

Our knowledge of global viral diversity in nature is still very incomplete and biased. Classical virology has focused heavily on viral pathogens infecting humans, livestock, and cultivars. The metagenomics revolution has begun to reveal a much broader picture, and suggests that we currently know <1% of the existing viruses in nature (25). Strategies to more efficiently discover viruses from new or underrepresented viral groups are needed. Currently, a large fraction of sequence reads (up to 90% in some cases) from environmental samples categorize as "dark matter," that is, sequences mapping to no other known biological entities (26). Remote homology detection tools and approaches that do not rely solely on comparative analysis of nucleic acid sequences are needed. For instance, prediction of viral capsid structures using artificial neural networks has recently contributed to identifying thousands of highly divergent new DNA viruses (27).

Characterizing the virosphere clearly goes beyond sequencing. We also need to address how different groups evolved from one another, how viral phylogenetic trees map to different ecosystems and hosts, why certain viral features have evolved recurrently in certain cases (e.g., multicomponent viruses mainly found in plants), how the ancestry of major viral groups is connected (e.g., whether insects served as evolutionary bridges between plant and animal viruses), and so on. The benefits of better characterizing and understanding the virosphere are manifold. For instance, we will achieve a broader picture of viral cross-species transmission, which should in turn inform viral emergence studies. We will also better identify factors involved in the transition from commensalism to parasitism.

IMPROVING OUR UNDERSTANDING OF VIRUS-VIRUS INTERACTIONS

The success of a viral infection sometimes depends not only on viral and host genotypes, but also on the presence of other microorganisms, including viruses, as revealed for instance when examining the viromes of flu-susceptible and flu-resistant wild birds (28). Some virus-virus interactions can be indirect, for instance by modifying the immune status of the host, whereas in other cases viruses interact more directly, such as in the case of satellite viruses. Virus-virus interactions can also involve members of the same viral species, which can cooperate or cheat according to the principles of social evolution (29). Well-known instances of such interactions include the emergence of defective viral genomes at high multiplicities of infection (30), but also more recently discovered processes such as cooperative immunity evasion in animal viruses and bacteriophages (31, 32). Moreover, viruses often undergo inter-host transmission or intra-host dissemination as pools of jointly dispersing viral particles, such as virion clusters enclosed in extracellular vesicles or virion aggregates (33), promoting coinfection. Whereas, virus-virus interactions typically take place when viruses share gene products in coinfected cells, it has been found that these interactions can also be established at the intercellular level. For instance, some temperate bacteriophages regulate lysis at the population level using small diffusible molecules encoded in their genomes, a system that resembles bacterial quorum sensing (34). Also, multicomponent viruses do not need to coinfect individual cells with particles containing each of the genome segments. Instead, viral products encoded by different segments are shared at the intercellular level in a process called distributed infection (35). The evolutionary stability of most of these processes depends on the spatial structure of viral infections, since this determines whether virus-virus interactions remain local or become systemic. As such, better understanding how viral infections progress anatomically and how this shapes spatial population structure remains another important research topic (36, 37).

At a broader scale, coinfections are at the origin of horizontal gene transfer among viruses, although this process can also be driven via gene transfer to the host and subsequent capture by other viruses. Either way, gene exchange plays a fundamental role in host range evolution, the virus-host arms race, and virus-host coevolution at large. Improving our knowledge of horizontal gene transfer in viruses is challenging for several reasons. For most viral taxa, the number of complete and closely related genomes is small, which affects the inference of gene exchange events. In more thoroughly studied groups, such as mycobacteriophages, the emerging picture is that extensive horizontal gene transfer has produced an evolutionary network of mosaic-like viral genome structures that is poorly described by classical dichotomous phylogenetic trees (38). However, due to the high substitution rates that characterize viral genomes, homologous genes exchanged in the past often diverge beyond the point where accurate inferences can be made. Finally, the high proportion of open reading frames with unknown function restricts our ability to understand the functional consequences of genome plasticity.

CONCLUSION

The above list of challenges is far from exhaustive, yet helps us discuss how certain active research areas might progress in the near future. Importantly, these challenges are clearly interconnected. For instance, improving systems for

experimental evolution should help us build more accurate predictions about viral evolution in nature, and better characterizing the virosphere should help us improve our understanding of virus-virus interactions. Some recent advances brought by these synergies have enforced deep changes in the virology and evolutionary biology communities, such as the way viral species are defined or the realization that most viruses are probably not pathogenic. Research at the interface between these areas should be encouraged too, such as for instance combining experimental virology and metagenomics. In the end, this should help us make viral evolution a more predictive, comprehensive and consistent scientific discipline.

AUTHOR CONTRIBUTIONS

All authors reviewed and approved the final version of the manuscript.

FUNDING

REFERENCES

1. Biek R, Pybus OG, Lloyd-Smith JO, Didelot X. Measurably evolving pathogens in the genomic era. *Trends Ecol. Evol.* (2015) 30:306–13. doi: 10.1016/j.tree.2015.03.009

2. Gouma S, Anderson EM, Hensley SE. Challenges of making effective influenza vaccines. *Annu Rev Virol.* (2020) 7:495–512. doi: 10.1146/annurev-virology-010320-044746

3. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. *Nature.* (2016) 530:228–32. doi: 10.1038/nature16996

4. Faria NR, Sabino EC, Nunes MRT, Alcantara LCJ, Loman NJ, Pybus OG. Mobile real-time surveillance of Zika virus in Brazil. *Genome Med.* (2016) 8:97. doi: 10.1186/s13073-016-0356-2

5. Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. *Euro Surveill.* (2017) 22:30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494

6. Rambaut A, Loman NJ, Pybus OG, Barclay W, Barrett J, Carabelli A, et al. *Preliminary Genomic Characterisation of an Emergent SARS-CoV-2 Lineage in the UK Defined by a Novel Set of Spike Mutations - SARS-CoV-2 Coronavirus/nCoV-2019 Genomic Epidemiology.* Virological (2020). Available online at: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563 (accessed March 8, 2021).

7. Dolan PT, Whitfield ZJ, Andino R. Mapping the evolutionary potential of RNA viruses. *Cell Host Microbe.* (2018) 23:435–46. doi: 10.1016/j.chom.2018.03.012

8. Gutierrez B, Escalera-Zamudio M, Pybus OG. Parallel molecular evolution and adaptation in viruses. *Curr Opin Virol.* (2019) 34:90–6. doi: 10.1016/j.coviro.2018.12.006

9. Zahradník J, Marciano S, Shemesh M, Zoler E, Chiaravalli J, Meyer B, et al. RBD *in vitro* evolution follows contagious mutation spread, yet generates an able infection inhibitor. *bioRxiv [preprint].* (2021). doi: 10.1101/2021.01.06.425392

10. Holmes EC, Rambaut A, Andersen KG. Pandemics: spend on surveillance, not prediction. *Nature.* (2018) 558:180–2. doi: 10.1038/d41586-018-05373-w

11. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. Host and viral traits predict zoonotic spillover from mammals. *Nature.* (2017) 546:646–50. doi: 10.1038/nature22975

12. Babayan SA, Orton RJ, Streicker DG. Predicting reservoir hosts and arthropod vectors from evolutionary signatures in RNA virus genomes. *Science.* (2018) 362:577–80. doi: 10.1126/science.aap9072

13. Han BA, Drake JM. Future directions in analytics for infectious disease intelligence: toward an integrated warning system for emerging pathogens. *EMBO Rep.* (2016) 17:785–9. doi: 10.15252/embr.201642534

14. Warren CJ, Sawyer SL. How host genetics dictates successful viral zoonosis. *PLoS Biol.* (2019) 17:e3000217. doi: 10.1371/journal.pbio.3000217

15. Hie B, Zhong ED, Berger B, Bryson B. Learning the language of viral evolution and escape. *Science.* (2021) 371:284–8. doi: 10.1126/science.abd7331

16. Moratorio G, Vignuzzi M. Monitoring and redirecting virus evolution. *PLoS Pathog.* (2018) 14:e1006979. doi: 10.1371/journal.ppat.1006979

17. Long A, Liti G, Luptak A, Tenaillon O. Elucidating the molecular architecture of adaptation via evolve and resequence experiments. *Nat Rev Genet.* (2015) 16:567–82. doi: 10.1038/nrg3937

18. Bertels F, Leemann C, Metzner KJ, Regoes R. Parallel evolution of HIV-1 in a long-term experiment. *Mol Biol Evol.* (2019) 36:2400–14. doi: 10.1093/molbev/msz155

19. Zinger T, Gelbart M, Miller D, Pennings PS, Stern A. Inferring population genetics parameters of evolving viruses using time-series data. *Virus Evol.* (2019) 5:vez011. doi: 10.1093/ve/vez011

20. Clevers H. COVID-19: organoids go viral. *Nat Rev Mol Cell Biol.* (2020) 21:355–6. doi: 10.1038/s41580-020-0258-4

21. Wagar LE, Salahudeen A, Constanz CM, Wendel BS, Lyons MM, Mallajosyula V, et al. Modeling human adaptive immune responses with tonsil organoids. *Nat Med.* (2021) 27:125–35. doi: 10.1038/s41591-020-01145-0

22. Ho SYW, Duchêne S, Molak M, Shapiro B. Time-dependent estimates of molecular evolutionary rates: evidence and causes. *Mol Ecol.* (2015) 24:6007–12. doi: 10.1111/mec.13450

23. Aiewsakun P, Katzourakis A. Time-dependent rate phenomenon in viruses. *J Virol.* (2016) 90:7184–95. doi: 10.1128/JVI.00593-16

24. Xue KS, Stevens-Ayers T, Campbell AP, Englund JA, Pergam SA, Boeckh M, et al. Parallel evolution of influenza across multiple spatiotemporal scales. *Elife.* (2017) 6:26875. doi: 10.7554/eLife.26875

25. Zhang Y-Z, Chen Y-M, Wang W, Qin X-C, Holmes EC. Expanding the RNA virosphere by unbiased metagenomics. *Annu Rev Virol.* (2019) 6:119–39. doi: 10.1146/annurev-virology-092818-015851

26. Krishnamurthy SR, Wang D. Origins and challenges of viral dark matter. *Virus Res.* (2017) 239:136–42. doi: 10.1016/j.virusres.2017.02.002

27. Tisza MJ, Pastrana DV, Welch NL, Stewart B, Peretti A, Starrett GJ, et al. Discovery of several thousand highly diverse circular DNA viruses. *Elife.* (2020) 9:51971. doi: 10.7554/eLife.51971.sa2

28. Wille M, Eden J-S, Shi M, Klaassen M, Hurt AC, Holmes EC. Virus-virus interactions and host ecology are associated with RNA virome structure in wild birds. *Mol Ecol.* (2018) 27:5263–78. doi: 10.1111/mec.14918

29. Dolgin E. The secret social lives of viruses. *Nature.* (2019) 570:290–2. doi: 10.1038/d41586-019-01880-6

30. Vignuzzi M, López CB. Defective viral genomes are key drivers of the virus-host interaction. *Nat Microbiol.* (2019) 4:1075–87. doi: 10.1038/s41564-019-0465-y

31. Borges AL, Zhang JY, Rollins MF, Osuna BA, Wiedenheft B, Bondy-Denomy J. Bacteriophage cooperation suppresses CRISPR-Cas3 and Cas9 immunity. *Cell.* (2018) 174:917–25. doi: 10.1016/j.cell.2018.06.013

32. Domingo-Calap P, Segredo-Otero EA, Durán-Moreno M, Sanjuán R. Social evolution of innate immunity evasion in a virus. *Nat Microbiol.* (2019) 4:1006–13. doi: 10.1038/s41564-019-0379-8

33. Sanjuán R, Thoulouze MI. Why viruses sometimes disperse in groups. *Virus Evol.* (2019) 5:vez01. doi: 10.1093/ve/vez025

34. Erez Z, Steinberger-Levy I, Shamir M, Doron S, Stokar-Avihail A, Peleg Y, et al. Communication between viruses guides lysis-lysogeny decisions. *Nature.* (2017) 541:488–93. doi: 10.1038/nature21049

35. Sicard A, Pirolles E, Gallet R, Vernerey M-S, Yvon M, Urbino C, et al. A multicellular way of life for a multipartite virus. *Elife.* (2019) 8:43599. doi: 10.7554/eLife.43599.018

36. Graw F, Perelson AS. Modeling viral spread. *Annu Rev Virol.* (2016) 3:555–72. doi: 10.1146/annurev-virology-110615-042249

37. Gallagher ME, Brooke CB, Ke R, Koelle K. Causes and consequences of spatial within-host viral spread. *Viruses.* (2018) 10:627. doi: 10.3390/v10110627

38. Hatfull GF. Mycobacteriophages. *Microbiol Spectr.* (2018) 6. doi: 10.1128/microbiolspec.GPP3-0026-2018