



## INFECTIOUS DISEASES

# The Global Virome Project

Expanded viral discovery can improve mitigation

By **Dennis Carroll, Peter Daszak, Nathan D. Wolfe, George F. Gao, Carlos M. Morel, Subhash Morzaria, Ariel Pablos-Méndez, Oyewale Tomori, Jonna A. K. Mazet**

Outbreaks of novel and deadly viruses highlight global vulnerability to emerging diseases, with many having massive health and economic impacts. Our adaptive toolkit—based largely on vaccines and therapeutics—is often ineffective because countermeasure development can be outpaced by the speed of novel viral emergence and spread. Following each outbreak, the public health community bemoans a lack of prescience, but after decades of reacting to each event with little focus on mitigation, we remain only marginally better protected against the next epidemic. Our ability to mitigate disease emergence is undermined by our poor understanding of the diversity and ecology of viral threats, and of the drivers of their emergence. We describe a Global Virome Project (GVP) aimed to launch in 2018 that will help identify the bulk of this viral threat and provide timely data for public health interventions against future pandemics.

Nearly all recent pandemics have a viral etiology with animal origins, and with their intrinsic capacity for interspecies transmission, viral zoonoses are prime candidates for

causing the next great pandemic (1, 2). However, if these viruses are our enemy, we do not yet know our enemy very well. Around 263 viruses from 25 viral families are known to infect humans (3) (see the figure), and given the rate of discovery following identification of the first human virus (yellow fever virus in 1901), it is likely many more will emerge in the future (4). We estimate, from analysis of recent viral discovery data (5), that ~1.67 million yet-to-be-discovered viral species from key zoonotic viral families exist in mammal and bird hosts—the most important reservoirs for viral zoonoses (supplementary text).

By analyzing all known viral-host relationships (3, 6), the history of viral zoonoses (7), and patterns of viral emergence (1), we can reasonably expect that between 631,000 and 827,000 of these unknown viruses have zoonotic potential (supplementary text). We have no readily available technological countermeasures to these as-yet-undiscovered viruses. Furthermore, the rate of zoonotic viral spillover into people is accelerating, mirroring the expansion of our global footprint and travel networks (1, 8), leading to a nonlinear rise in pandemic risk and an exponential growth in their economic impacts (8).

## PROMISING PILOT, CHALLENGING SCALE

Since 2009, the U.S. Agency for International Development (USAID) has conducted a large-scale pilot project, spanning more than 35 countries over 8 years at a cost of around \$170 million, to evaluate the feasibility of preemptively mitigating pandemic threats.

Scientists prepare to collect a blood sample from a *Rousettus sp.* fruit bat in Thailand to test for novel viruses. The Global Virome Project aims to identify and characterize the majority of currently unknown viruses in key wildlife groups, including rodents, nonhuman primates, and bats.

Other previous studies had begun to conduct targeted viral discovery in wildlife (9), and develop mitigation strategies for the emergence of avian flu, for example. However, the USAID Emerging Pandemic Threats (EPT) PREDICT project is the first global-scale coordinated program designed to conduct viral discovery in wildlife reservoir hosts, and characterize ecological and socioeconomic factors that drive their risk of spillover, to mitigate their emergence in people (10).

Working with local partners and governments, wildlife and domestic animals and at-risk human populations in geographic hotspots of disease emergence (1) are sampled, and viral discovery conducted. A strategy to identify which novel viruses are most at risk of spillover has been developed (11), and further work is conducted on these to characterize them prior to, or in the early stages of, spillover. Metadata on the ecology of wildlife–livestock–human transmission interfaces, and on human behavioral patterns in communities, are concurrently analyzed so that strategies to reduce spillover can be developed (supplementary text). To date, EPT PREDICT has discovered more than 1000 viruses from viral families that contain zoonoses, including viruses involved in recent outbreaks (12), and others of ongoing public health concern (13). The focus of EPT PREDICT on capacity building, infrastructure support, training, and epidemiological analysis differs substantially from the GVP's emphasis on large-scale sampling and viral discovery. However, to discover the bulk of the projected remaining 1.67 million unknown viruses in animal reservoirs and characterize the majority of 631,000 to 827,000 viruses of highest zoonotic potential requires overcoming some challenges of scale.

The first challenge is cost. To estimate this, we analyzed data on field sampling and laboratory expenditures for viral discovery from (5, 10), and estimates of unknown viral diversity in mammalian and avian hosts (supplementary text). We estimate that discovery of all viral threats and characterization of their risk for spillover, using currently available technologies and protocols, would be extremely costly at over \$7 billion (supplementary text). However, previous work shows that viral discovery rates are vastly higher in the early stages of a sampling program, and that discovering the last few, rare, viruses is extremely costly and time-consuming owing to the number of samples required to find

The list of author affiliations is provided in the supplementary materials. Email: [daszak@ecohealthalliance.org](mailto:daszak@ecohealthalliance.org)

them (5) (supplementary text). We used data on rates of viral discovery (5) to estimate that the substantial majority of the viral diversity from our target zoonotic reservoirs could be discovered, characterized, and assessed for viral ecology within a 10-year time frame for ~\$1.2 billion (16% of total costs for 71% of the virome, considering some fixed costs) (fig. S1). Those viruses remaining undiscovered will, by the nature of sampling bias toward more common host species, represent the rarest viruses with least opportunity for spillover, and therefore reduced public health risk. Their discovery would require exponentially greater sampling effort and funding that could be better spent on countermeasures for the more likely threats (supplementary text).

Stakeholders from Asia, Africa, the Americas, and Europe, spanning industry, academia, intergovernmental agencies, non-governmental organizations (NGOs), and the private sector, began meeting in 2016 to design a framework for the governance, management, technical operation, and scope of the GVP. Key efforts include developing finance streams; establishing a transparent, equitable implementation strategy; designing data- and sample-sharing protocols; developing laboratory and metadata platforms; targeting of host taxa and sampling sites; analyzing return on investment; forming collaborative field and laboratory networks; developing risk characterization frameworks for viruses discovered; designing a strategy to assess and mitigate risk behaviors that facilitate viral emergence; and planning in-country capacity building for sustainable threat mitigation. Funding has been identified to support an initial administrative hub, and fieldwork is planned to begin in the first two countries, China and Thailand, during 2018.

With outputs intended to serve the global public good, the GVP is developing a transparent and equitable strategy to share data, viral samples, and their likely products, including benefits derived from future development of medical countermeasures. These build on the Nagoya Protocol to the Convention on Biological Diversity and the Pandemic Influenza Preparedness Framework, negotiated by the World Health Organization (WHO). The international collaborative nature and global ownership of the project should help leverage funding from diverse international donors, including government agencies focused on national virome projects or on international development projects in other countries, and private-sector philanthropic donors focused on technology and big science.

The diversity of tasks required to conduct the GVP should reduce the potential for it to divert funds from current public health programs. For example, discrete work streams on

targeted sampling of wildlife, on bioinformatics, and on behavioral risk analysis fall within the focus of current scientific research programs in a range of donor agencies. Governments and corporations with specific remits and geographic responsibilities have been approached to finance subprojects relevant to their sectors (e.g., capacity development, surveillance of specific taxa, geographically focused activities, medical countermeasure development, training, surveillance, and technological platforms). In addition, leaders in China and a number of countries have begun developing national virome projects as part of the GVP, leveraging current research funding to include GVP sites.

Technological challenges include safe field sampling in remote locations and cost-effective

---

***“...the GVP goals...improve capacity to detect, diagnose, and discover viruses in vulnerable populations...”***

laboratory platforms that can be standardized in low-income settings. To achieve these goals, existing national, regional, and international networks will need to be enhanced and expanded within standardized sampling and testing frameworks. Existing networks of field biologists from environment ministries, academic institutions, and conservation and health NGOs may assist in surveillance. National science and technology agencies, regional One Health platforms, transboundary disease surveillance networks, Institut Pasteur laboratories, WHO, United Nations Food and Agricultural Organization, and the World Organization for Animal Health collaborating, and reference centers and viral discovery laboratories, including USAID EPT PREDICT, are currently involved in planning these activities around a decade-long sampling and testing time frame. A monitoring and evaluation strategy is being developed based on analysis of viral discovery rates against predicted viral diversity, to identify when to halt surveillance and testing as the GVP progresses. Stakeholders will also tackle the challenge of how to decide when enough potentially dangerous viruses have been discovered in a host species or region to call for action to reduce underlying drivers of emergence (e.g., hunting and trading of a wildlife reservoir).

Laboratory platforms developed by USAID EPT PREDICT have proven capacity to identify novel viruses and are relatively inexpensive and reliable, being based on polymerase chain reaction using degenerate

primers that target a range of viral families of known zoonotic potential. However, scaling up to a full global virome project will require discovery of three orders of magnitude more viruses in a similar time frame. Technological solutions will be needed to increase the speed and efficiency, and reduce the cost, of sequence generation. These will likely include next-generation sequencing and other unbiased approaches to identify evolutionarily distinct viral clades.

A key challenge is how to assess the potential for novel viruses to infect people or become pandemic (14). The EPT PREDICT project (11) and others (2, 6) have developed preliminary zoonotic risk characterization frameworks based on viral and host traits and the ecological and demographic characteristics of the sampling site. These approaches will be used in the GVP to triage novel viruses for further characterization to assess their zoonotic capacity (supplementary text). In vitro receptor binding analyses coupled with in vivo models have proven useful in this capacity for some viral families [e.g., coronaviruses (13)]. Although this is not yet feasible for all potentially zoonotic viral clades, applying these techniques to a larger viral data set as the GVP progresses will allow validation of risk frameworks and may increase our capacity to predict zoonotic potential. However, advancing these goals will require new collaboration among lab virologists, epidemiologists, and modelers, innovative approaches to field-testing the boundaries of virus-host relationships, and support across agencies that often fund separate virology, public health, evolutionary biology, and biodiversity modeling initiatives.

#### **INVESTMENTS, RETURNS**

The cost of the GVP represents a sizable investment and, even if a large number of potential zoonoses are discovered, only a minority is likely to have the potential to cause large-scale outbreaks and mortality in people (1, 2, 7). However, given the high cost of single epidemic events, data produced by the GVP may provide substantial return on investment by enhancing diagnostic capacity in the early stages of a new disease outbreak or by rapidly identifying spillover hosts, for example. Recent analysis of the exponentially rising economic damages from increasing rates of zoonotic disease emergence suggests that strategies to mitigate pandemics would provide a 10:1 return on investment (1, 8). Even small reductions in the estimated costs of a future influenza pandemic (hundreds of billions of dollars) or of the previous SARS (severe acute respiratory syndrome) epidemic (\$10 billion to \$30 billion) could be substantial. The goal of the GVP is to improve efficiency in the face of these increasing viral



spillover rates by enhancing (not replacing) current pandemic surveillance, prevention, and control strategies. If we were to invest only in surveillance for known pathogens (our current business-as-usual strategy), our calculations suggest we would protect ourselves against less than 0.1% of those viruses that could conceivably infect people, even using the lower bounds of our uncertainty for our viral estimates (i.e., 263 viruses known from humans out of 263,824 unknown potential zoonoses; supplementary text).

The potential benefits of the GVP may be enhanced to maximize public health benefits (supplementary text) by (i) optimizing sampling to target species most likely to harbor “missing zoonoses” (6), or to target emerging disease hotspot regions most likely to propagate major disease outbreaks (7); (ii) using human and livestock syndromic surveillance to identify regions for wildlife sampling proximal to repeated outbreaks of severe influenza-like-illnesses, fevers of unknown origin, encephalitides, livestock “abortion storms,” and other potential emerging disease events; (iii) initially targeting RNA viruses, which caused 94% of the zoonoses documented from 1990 to 2010; and (iv) fostering economies of scale and adoption of technological innovation as the GVP ramps up. This includes use of laboratories that can facilitate regional sample processing, development of centralized bioinformatics platforms, and improved logistics for sample collection and transport. We also expect the cost of testing and sequencing to decrease as technology is enhanced, much as the development of next-generation sequencing reduced genetic sequencing costs by up to four orders of magnitude in a decade.

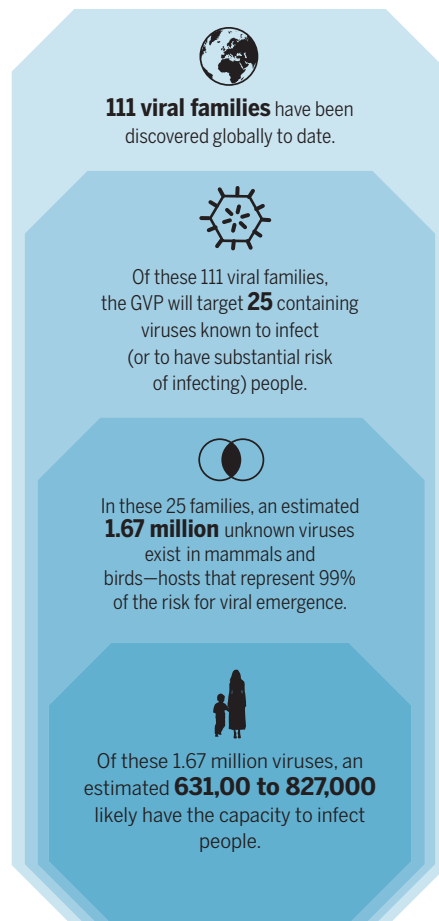
The accelerated pace of viral discovery under the GVP will make the virological, phylogenetic, and modeling approaches used in pandemic preparedness more data-rich, and likely more effective. For example, having the sequence data for thousands, rather than a few, viruses from a single family could extend vaccine, therapeutic, or drug development to a wider range of targets, leading to broad-spectrum vaccines and other countermeasures. Identification of novel viruses may be useful to programs like the Coalition for Epidemic Preparedness Innovations (CEPI) in assessing the breadth of action of candidate vaccines and therapeutics, and in expanding their efficacy. More broad-scale prevention approaches could provide immediate return on investment prior to vaccine and countermeasure development, which would require substantial investment and time. For example, metadata on viral reservoir host identity, geography, seasonality, proximity to people, and drivers of emergence will refine our mechanistic understanding of spillover

and enhance published models of emerging infectious diseases risk (1, 6). Identification of novel viruses in hunted, traded, or farmed wildlife species could be used to enhance biosecurity in markets and farming systems, reducing public health risk, increasing food security, and assisting in conservation of hunted species. The presence of hosts harboring high-risk novel viruses in proximity to human populations may allow targeted follow-up to examine evidence of spillover and design intervention strategies (supplementary text). Ultimately, the benefits of the GVP may include enhancing our understanding of viral biology, such as drivers of competition or cooperation among viruses within hosts, genomic underpinnings of host-virus coevolution, processes underlying deep evolution of viral clades, and the identification of novel viral groups (15).

The regions targeted by the GVP are largely highly biodiverse, rapidly developing countries in the tropics, which often have low

### GVP targeting strategy

The project will capitalize on economies of scale in viral testing, systematically sampling mammals and birds to identify currently unknown, potentially zoonotic viruses that they carry.



capacity to deal with public health crises (1). The expanded laboratory capacity, field sampling, and data generation intrinsic to the GVP goals will therefore improve capacity to detect, diagnose, and discover viruses in vulnerable populations within regions most critical to preventing future pandemics. This enhanced capacity may also help improve diagnosis and control for endemic diseases, as well as the portion of the virome that remains undiscovered.

The Human Genome Project in the 1980s catalyzed technological innovation that dramatically shortened the time and cost for its completion, and ushered in the era of personalized genomics and precision medicine. The GVP will likely accelerate development of pathogen discovery technology, diagnostic tests, and science-based mitigation strategies, which may also provide unexpected benefits. Like the Human Genome Project, the GVP will provide a wealth of publicly accessible data, potentially leading to discoveries that are hard to anticipate, perhaps viruses that cause cancers and chronic physiological, mental health, or behavioral disorders. It will provide orders-of-magnitude more information about future threats to global health and biosecurity, improve our ability to identify vulnerable populations, and enable us to more precisely target mitigation and control measures to foster an era of global pandemic prevention. ■

### REFERENCES AND NOTES

1. K. E. Jones *et al.*, *Nature* **451**, 990 (2008).
2. J. L. Geoghegan, A. M. Senior, F. Di Giallonardo, E. C. Holmes, *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4170 (2016).
3. A. M. King, M. J. Adams, E. B. Carstens, E. J. Lefkowitz, *Virus Taxonomy: Classification and Nomenclature of Viruses: Ninth Report of the International Committee on Taxonomy of Viruses* (Elsevier, Philadelphia, 2012).
4. M. Woolhouse, F. Scott, Z. Hudson, R. Howey, M. Chase-Topping, *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2864 (2012).
5. S. J. Anthony *et al.*, *mBio* **4**, e00598-13 (2013).
6. K. J. Olival *et al.*, *Nature* **546**, 646 (2017).
7. N. D. Wolfe, C. P. Dunavan, J. Diamond, *Nature* **447**, 279 (2007).
8. J. Pike, T. L. Bogich, S. Elwood, D. C. Finnoff, P. Daszak, *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18519 (2014).
9. J. F. Drexler *et al.*, *Nat. Commun.* **3**, 796 (2012).
10. data.predict.global
11. S. S. Morse *et al.*, *Lancet* **380**, 1956 (2012).
12. G. Grard *et al.*, *PLOS Pathogens* **8**, e1002924 (2012).
13. X.-Y. Ge *et al.*, *Nature* **503**, 535 (2013).
14. J. L. Geoghegan, E. C. Holmes, *Open Biol.* **7**, 170189 (2017).
15. C. X. Li *et al.*, *eLife* **4**, e05378 (2015).

### ACKNOWLEDGMENTS

P.D., N.D.W., and J.A.K.M. are funded by USAID EPT PREDICT. We acknowledge S. J. Anthony, C. J. Chrisman, Y. Feferholtz, T. Goldstein, C. K. Johnson, D. Nabarro, K. J. Olival, N. Ross, E. Rubin, R. Waldman, B. Watson, C. Zambrana-Torrel, attendees of the Rockefeller Foundation-funded Bellagio Center Global Virome Project Workshop August 2016 ([www.globalviromeproject.org/about/](http://www.globalviromeproject.org/about/)), members of the GVP Core Group and Steering Committee, and co-leads of the GVP Working Groups for their help refining the concept and this manuscript.

### SUPPLEMENTARY MATERIALS

[www.sciencemag.org/content/359/6378/872/suppl/DC1](http://www.sciencemag.org/content/359/6378/872/suppl/DC1)

10.1126/science.aap7463

## The Global Virome Project

Dennis Carroll, Peter Daszak, Nathan D. Wolfe, George F. Gao, Carlos M. Morel, Subhash Morzaria, Ariel Pablos-Méndez, Oyewale Tomori and Jonna A. K. Mazet

*Science* **359** (6378), 872-874.  
DOI: 10.1126/science.aap7463

### ARTICLE TOOLS

<http://science.sciencemag.org/content/359/6378/872>

### SUPPLEMENTARY MATERIALS

<http://science.sciencemag.org/content/suppl/2018/02/21/359.6378.872.DC1>

### REFERENCES

This article cites 13 articles, 5 of which you can access for free  
<http://science.sciencemag.org/content/359/6378/872#BIBL>

### PERMISSIONS

<http://www.sciencemag.org/help/reprints-and-permissions>

Use of this article is subject to the [Terms of Service](#)

---

*Science* (print ISSN 0036-8075; online ISSN 1095-9203) is published by the American Association for the Advancement of Science, 1200 New York Avenue NW, Washington, DC 20005. 2017 © The Authors, some rights reserved; exclusive licensee American Association for the Advancement of Science. No claim to original U.S. Government Works. The title *Science* is a registered trademark of AAAS.