www.sciencemag.org/content/359/6378/872/suppl/DC1

Supplementary Materials for

The Global Virome Project

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

*Corresponding author. Email: daszak@ecohealthalliance.org

Published 23 February 2018, *Science* **359**, 872 (2018)
DOI: 10.1126/science.aap7463

This PDF file includes:

List of Author Affiliations

Dennis Carroll[1], Peter Daszak[2*], Nathan D. Wolfe[3], George F. Gao[4,5], Carlos M. Morel[6], Subhash Morzaria[7], Ariel Pablos-Méndez[8], Oyewale Tomori[9], Jonna A.K. Mazet[10]

[1]Emerging Threats Division, Bureau for Global Health, U.S. Agency for International Development, Washington, District of Columbia, U.S.A.

[2]EcoHealth Alliance, 460 West 34th Street – 17th Floor, New York, New York, U.S.A.

[3]Metabiota, 425 California Street – 2nd Floor, San Francisco, California, USA.

[4]CAS Key Laboratory of Pathogenic Microbiology and Immunology, Institute of Microbiology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, People's Republic of China.

[5]Chinese Center for Disease Control and Prevention (China CDC), 155 Changbai Road, Changping District, Beijing 102206, People's Republic of China

[6]National Institute of Science and Technology for Innovation in Neglected Diseases (INCT-IDN), Centre for Technological Development in Health (CDTS), Oswaldo Cruz Foundation (Fiocruz), Av. Brasil 4365, Rio de Janeiro, RJ 21040-900, Brazil.

[7]Animal Health Service, Food and Agriculture Organization, United Nations, Viale delle Terme di Caracalla, 00153 Rome, Italy.

[8]Columbia University Medical Center, 622 West 168th Street, PH-9E-103c, New York, New York, U.S.A.

[9]The World Academy of Sciences & Nigerian Academy of Sciences, Academy House, 8A Ransome Kuti Road, University of Lagos, Lagos, Nigeria.

[10]One Health Institute, School of Veterinary Medicine, University of California, Davis, California, USA.

*Correspondence to: daszak@ecohealthalliance.org

Supplementary Text

The steps taken to estimate the global number of mammalian and avian viruses are detailed below. Data and code necessary to reproduce the analysis are available at **http://www.github.com/ecohealthalliance/GVP_Science**.

Estimating viral diversity in wildlife hosts of zoonotic relevance

To estimate the diversity of viruses that threaten to emerge in people we extrapolated from estimates derived from two mammalian wildlife hosts, a bat species and a primate species, representing the only published estimates of unknown viral diversity in vertebrate hosts *(7, 16)*. Mammals represent the most important hosts for emerging viral zoonoses, with ~88% (69/78) of emerging viral zoonoses with a confirmed reservoir host in a published database of emerging infectious disease events originating in non-human mammals (*1*), and ~99% in non-human mammals or in birds (69 from mammals, 8 from birds). These include viruses that originate in, or have genes that originate in, these classes of vertebrates or that are transmitted by them at some stage in their emergence. In many biological studies, estimates of diversity include the number of species or other taxonomic units in a region (the species richness), as well as the variability within species, among species, and across landscapes, i.e. the degree of difference among organisms (*17*). For the purposes of this paper, we follow previous work (*5, 16*) and use the phrase 'viral diversity' to denote the species richness of viruses, i.e. the number of species of viruses in a given host, group of hosts or geographic region. Estimates of viral diversity for the bat *Pteropus giganteus* are based on repeated sampling of individuals of this species and viral discovery using PCR with degenerate family-level primers targeting 9 viral genera or families, all known to contain zoonoses (*5*). Estimates of viral diversity for the rhesus macaque *Macaca mulatta* targeted 21 viral families by PCR, as well as high-throughput sequencing (HTS) (*16*). For each host species, we obtained raw test data from (*5, 16*) ((https://github.com/ecohealthalliance/GVP_Science/blob/master/data/macaque_data.xlsx), (https://github.com/ecohealthalliance/GVP_Science/blob/master/data/pteropus_data.csv)) and repeated the analysis of unknown viral diversity for each set of viral family-level tests to estimate the expected number of viral species and standard error for all known and expected unknown viruses from each of the 30 viral family test protocols conducted (Code available at https://github.com/ecohealthalliance/GVP_Science/blob/master/R/pteropus_validation_anthony_2013.R and https://github.com/ecohealthalliance/GVP_Science/blob/master/R/macaque_validation_anthony_2015.R, respectively). This gave an average of 11.58 viruses per family (standard error 6.74), indicating that the average viral family has between 4.84 and 18.32 viral species. However, this mean was heavily skewed by the 120 discovered and 200.83 estimated Picobirnoviruses in *M. mulatta*. The median number of viral species per viral family was 2.5, indicating that many viral families have only a few viral species in a given host, but rare families can have many more. For the purposes of the Global Virome Project (GVP), the mean is a more useful metric so that costs are not underestimated. The likely variability in viral diversity among mammalian species and viral families highlights the need for a strategy to monitor sampling and test results in real time to guide per-species sampling as the GVP proceeds (see below).

To decide which viral families to include in GVP testing protocols, we used the stringent database from (*6*) that requires PCR evidence of viral sharing among humans and other animals, and is derived from (*3*). All 21 viral families targeted in (*5, 16*) contain known zoonoses, based on PCR reports in the database from (*6*), but neither study covered the Hepeviridae, which also contains zoonoses, so this was included. The Arteriviridae contains a species pathogenic in primates and none that are known from humans, but could reasonably be considered a risk for future zoonotic viral emergence from either known or unknown viral *Arterivirus* species, given close evolutionary relationships between humans and other primates. The Bornaviridae was not included because the evidence for its zoonotic origin is unclear, and most reports of animal

infections are from livestock, not wildlife (*18, 19*). The Hepadnaviridae was included because there is PCR and *in vitro* infection evidence of viral origin in wildlife species that are known high zoonotic risk reservoirs, suggesting high likelihood of related novel viruses capable of spilling over into people (*20*). We therefore multiply the per-viral family mean of 11.58 species by 25 (see specific viral families listed below) to give an estimate of the total viral diversity within all 25 viral families that have high zoonotic potential as 289.5 (se 121.0 - 458.0) for each mammalian host species. Given that there are 5,291 known mammal species (*21*) (excluding marine mammals which have limited contact with humans relative to terrestrial and volant mammals) we estimate global mammalian viral richness in our 25 target viral families at 1,531,745 (se 640,211 - 2,423,278). In addition, influenza viruses derived from mammals and birds represent a significant public health threat globally. Therefore, we used the mean viral diversity per viral family in mammal hosts (11.58), and assumed that this may also represent the average per-host species viral diversity for influenza viruses (Orthomyxoviridae) in each of the 11,862 known bird species bird species (*22*). This gives an estimated avian influenza viral richness of 137,362.96 (se 57,412 - 217,312), and a total global richness of viral species from families with high zoonotic potential of around 1,669,106 (se 697,623 - 2,640,590).

The viral families to be targeted by the GVP include 16 families of RNA viruses known to infect people: Arenaviridae, Astroviridae, Bunyaviridae, Caliciviridae, Coronaviridae, Filoviridae, Flaviviridae, Hepeviridae, Orthomyxoviridae, Paramyxoviridae, Picobirnaviridae, Picornaviridae, Reoviridae, Retroviridae, Rhabdoviridae, and Togaviridae; an additional RNA viral family (Arteriviridae) that infects non-human primates will be included, as explained above; and an additional 8 families of DNA viruses which contain known zoonoses (Adenoviridae, Anelloviridae, Hepadnaviridae, Herpesviridae, Papillomaviridae, Parvoviridae, Polyomaviridae, Poxviridae), bringing the total to 25 viral families.

There are a number of assumptions in this approach that need to be considered. Firstly, viruses, and RNA viruses in particular, have remarkable potential for evolution which suggests that estimates of viral diversity may not remain static over the period of the GVP. However, the large standard error around the GVP estimates of viral diversity suggests that they err on the side of over-estimation of viral diversity by using the mean per-family value. Therefore, the evolution of novel viral lineages during the GVP is unlikely to significantly alter total diversity discovered. Furthermore, the work on which these estimates are based used a conservative approach to the definition of a viral species, and this has been followed in the current calculations (*5, 16*). It is also likely that viral evolution is more constrained in reservoir hosts (the subject of the GVP sampling and testing) than in spillover hosts (e.g. humans, the subject of many studies of RNA viral evolution, e.g. (*23*), which may further reduce the impact of viral evolution on these estimates. Secondly, the estimates of viral diversity are based on two species, the macaque and the giant fruit bat, which may not be representative of all mammalian or avian species in that they are both mammals, both social, the former is widely distributed and the latter lives in dense colonies and is volant. These and other factors (e.g. overlapping host species ranges resulting in viral sharing, long evolutionary distance

from related host species resulting in diversification of viral lineages) suggest that our estimates of global viral diversity may be lower or higher than stated. Some of these have been discussed in (*5, 16*), and given the large standard errors presented here for our estimates of viral diversity, seem unlikely to suggest a significant disparity (e.g. by orders of magnitude) between the estimated viral diversity and that which would eventually be discovered. However, to deal with potential disparities, we propose that the testing results from the GVP should be analyzed routinely as the project progresses, to monitor and evaluate progress relative to initial estimates and assumptions. Specifically, the same algorithms used in (*5, 16*) could be applied to viral family test results per host species, family, order, and geographical sampling region to ascertain whether the standard error around estimated total viral diversity reduces as the project progresses. A reduction in these standard errors would denote a gain in accuracy of estimates, and if these suggest some viral families are far more (or less) speciose than assumed, or some species or taxa have a much greater or lesser viral species richness than assumed, the strategic targeting of the GVP could be altered accordingly to maximize discovery, by expanding, reducing or halting sampling and testing.

Estimating the number of potentially zoonotic unknown viruses

To estimate the number of wild mammal viruses and bird influenzas with the potential to infect people, (i.e. the species richness of unknown viral zoonoses), we used data from (*3, 6*) on the host species range for all known mammalian viruses and the proportion of those viruses that can infect humans. We found that, of 580 known mammalian viruses in the 25 viral families targeted by the GVP, 261 (45.0%) infect humans (2 other viruses are not within these families: the unassigned Hepatitis delta virus, and Borna disease virus-1; bringing the total known human viruses from (*3, 6*) to 263). (https://github.com/ecohealthalliance/HP3/blob/master/data/viruses.csv) As evidence for a mammalian viral species' capacity to infect people, we used data from the less stringent database in (*6*), which includes those with either PCR or serology evidence of infection. This reduces the likelihood of underestimating zoonotic potential, and reflects the large number of studies of spillover potential that use serological surveys of exposed people. We did not include Hepatitis delta virus, in our calculations because this is considered an RNA viroid, requires hepatitis B virus (HBV) as a 'helper virus' for its replication and transmission, and is only known in nature from HBV-infected humans (*24-26*). Furthermore, it has not been categorized into a viral family by the ICTV. Of the 580 known mammalian viruses in the 25 viral families targeted by the GVP, 74 are thought to be exclusively human and 187 are zoonotic (32.2% of the known mammalian viruses in the targeted families). Considering evidence that most exclusively human viruses appear to have originated in animals (*7*), we assume that between 493,856 (se 206,412 - 781,298) (32.2%) and 689,285 (se 288,095 - 1,090,475) (45.0%) of the total unknown mammalian viruses have potential to infect humans. Because influenza viruses have exceptional capacity for genetic recombination, we made the conservative assumption that all strains of influenza virus in birds are likely capable of zoonotic transmission. Our best estimate of the total number of potential zoonoses globally is therefore between 631,218 (se 263,824 – 998,610) and 826,647 (se 345,507 - 1,307,787). It is possible that as-yet-undiscovered viruses harbor a lower proportion with zoonotic potential than those already known, given our long history of interaction with wildlife on

the planet, and our likely skewed surveillance of hosts known to harbor zoonoses (*6*). However, data show that the rate of disease emergence is increasing after correcting for reporting bias, with zoonotic pathogens originating in wildlife increasing significantly as a proportion of all emerging pathogens (*1, 8*). This suggests that any reduced potential for zoonotic emergence in the undiscovered pool may be far less important in driving risk than the very large number of pathogens to which humans have not yet been significantly exposed, or the exponentially increasing drivers of high-risk contact with wildlife that leads to disease emergence (*1, 27*). Furthermore, the identification of a pathogen as zoonotic requires that it infects people and is then clinically observed. Capacity to conduct disease detection and viral discovery in people, livestock and wildlife is likely lowest in countries with the highest mammalian biodiversity (and therefore viral diversity), so that many viruses may remain undiscovered, even if they have already emerged repeatedly into human populations in these regions. This hypothesis is supported by the apparent repeated spillover of HIV-1 progenitors prior to its emergence (*28*), and PCR evidence of simian foamy virus infections in three bushmeat hunters in Cameroon in a survey of 1099 individuals, ten of which had serological evidence of exposure (*29*). In addition, analysis of ICTV data and the literature (*6*) suggests that prior surveillance studies have not skewed to hosts with higher proportions of zoonoses due to research focused on identifying reservoirs of zoonotic pathogens. For example, this work shows that bats have a statistically significant higher proportion of zoonoses than any other mammalian order, even though bat viruses are not historically well-studied relative to other mammalian host groups (*21*). Indeed, other factors may determine the selection of hosts for surveillance, not least of all convenience. For example, historical surveys of mammals are often based on ease of capture, and therefore have focused more on diurnal terrestrial species (e.g. rodents, primates) than nocturnal volant species (e.g. bats). Finally, the standard error for our estimates of the number of viruses with zoonotic potential is quite large, but still results in a minimum expectation of >220,000 viruses of zoonotic potential. Given the high, and growing, public health impact and economic costs of EID outbreaks (*8*), it seems that this lower number would still provide adequate economic and public health rationale for a large project like the GVP.

Targeting the Global Virome Project to maximize discovery of zoonoses

We propose that work conducted by the Global Virome Project (GVP) should be streamlined to discover those viruses with the highest zoonotic potential by targeting host taxa which have historically harbored the majority of zoonoses. Mammalian taxa sampled could reasonably exclude marine mammals, which have a reduced likelihood of contact with people relative to terrestrial and volant mammals, and have not been the primary reservoir from which any prior EIDs have originated (*1*). Target mammalian species primarily would be wildlife because 91% of zoonoses reported between 1990 and 2010 involved a wildlife host, with some of those also implicating a domestic animal host for the same virus etiology (*30*). For avian influenza viruses, the GVP could reasonably target only water birds (Anatidae, Laridae, Charadriiformes and others, with a global species richness of 871) which harbor a wide diversity of known zoonotic avian influenzas, and include species that are closely related to widely-domesticated species, and that are hunted extensively (*22*). Some recent emerging diseases have been caused by viruses transmitted to people from wildlife via domestic animal reservoirs (e.g. Nipah

virus and pigs, MERS coronavirus and camels). The GVP could sample a small number of livestock and poultry species (n=10) in regions that are known emerging disease hotspots (*1, 27*). Because these species are abundant and widely distributed, sampling in 5 regions globally would not be logistically difficult, and would increase the target by an equivalent of 50 species.

Estimating the cost of a Global Virome Project

To estimate the cost of sampling and discovering unknown viruses, we used cost data derived from the USAID EPT/PREDICT project, which has a goal of conducting zoonotic viral discovery in wildlife to identify potentially zoonotic pathogens and build capacity to reduce the risk of pandemics (*10, 11*). This project has been operating in over 25 developing countries, in emerging disease hotspots, for just over 8 years. This project has set up field teams in each country, captured and sampled over 80,000 wildlife and domestic animals, designed and operated diagnostic platforms in-country, and conducted over 400,000 viral discovery tests. The USAID EPT/PREDICT project conducted the work in (*5, 16*). The cost of estimating the viral diversity of *P. giganteus*, collecting metadata, conducting risk assessment and partially characterizing some of the viruses discovered was $1.2 million. The bulk of these costs were in sample collection, setting up laboratory testing platforms, and conducting sample preparation. Considering that the cost of adding primers to PCR reactions once samples are prepared for analysis, including HTS, is relatively insignificant, we estimate the cost of taking the Global Virome Project to scale is approximately $6.3492 billion (the $1.2 million cost of the *P. giganteus* work multiplied by the number of extant terrestrial and volant mammals) for 100% of the estimated unknown 1,531,745 mammalian viruses (excluding marine mammals). Because the bulk of these costs comprise the logistics of fieldwork, sample collection and cold chain management, they are therefore driven largely by host species considerations and relatively unaffected by the uncertainty in viral diversity estimates. The GVP would require an additional $1.1052 billion to sample water bird species (n=871) for influenza, and to conduct limited domestic animal sampling (n=50, as described in the SM, above). This would bring the total costs of the GVP to $7.454 billion, which, if annualized over a decade would cost around $745 million per year to discover virtually 100% of the viral zoonotic threats to our species.
(https://github.com/ecohealthalliance/GVP_Science/blob/master/data/GVP_costs_from_anthony2013.xlsx)

However, because the viral discovery curves reported in (*5, 16*) are asymptotic, marginal returns (i.e. rates of new virus discovery) diminish rapidly as the number of samples collected increases. Using the discovery rates and total costs from (*5*), and assuming a linear relationship between samples collected, we examined the relationship between total project costs and the percentage of the undiscovered global virome found. (Fig. S1). Based on these calculations, we estimate that 97.4% of targeted viral diversity would be discovered with half of the sampling and testing effort ($3.73 billion), 85% of targeted viral diversity would be discovered with only 23% of the costs ($1.69 billion), and 71% diversity with only 16% of the costs ($1.2 billion, considering some fixed costs). Neither start-up infrastructure costs nor cost savings from economies of scale are included in this model, both of which may reduce the steepness of the curve shown in Fig. S1, but this simple model is included to demonstrate the decreasing rate of discovery after an initial sampling phase. This model suggests that a GVP costing $1.2 billion could discover a substantial majority (71%) of unknown viral diversity in non-marine mammals and of the zoonotic influenza risk from water birds, with some limited domestic animal sampling. GVP costs likely would not be significantly increased if estimates of viral diversity prove to be low due to, for example, the discovery of a very diverse new viral family, or greater diversity within a known family. The increased number of viruses discovered would not be linearly translated into increased costs of the project, because

the bulk of the costs are in the collection of field samples rather than sequencing a larger number of PCR products.

Further strategic targeting to increase return-on-investment

Further targeting strategies are proposed in the main text to increase return-on-investment from the GVP, and deliver public health value as rapidly as possible: *1) Optimizing sampling to target host species and regions with the highest number of 'missing zoonoses', or hosts otherwise more likely to be involved in viral spillover.* It is logical that some host species, genera or orders may have a higher risk than others of harboring zoonoses, or of their viral fauna spilling over to people. This may be due to their higher contact with people (e.g. through a propensity to live in human-dominated habitat, or their involvement in the wildlife trade etc.); due to intrinsic biological factors that lead to them harboring a higher number of viruses, or likely zoonotic viruses; or because they are more abundant or have a wider geographic or ecosystem range, and therefore increased potential for direct viral spillover or via domestic animals to humans. Recent analyses suggest that species which have greater contact with humans (*31*), and are more closely related phylogenetically (*32, 33*) are more likely to harbor viruses with the potential to be zoonotic. Analysis of data on all mammalian host-virus relationships in (*3*), after controlling for reporting effort, demonstrates that both the total number of viruses that infect a given species, and the proportion likely to be zoonotic are predictable (*6*). The proportion of viral species that are zoonotic in mammalian hosts is predicted by phylogenetic relatedness to humans, host taxonomy, and human population within a species range (a proxy for human-wildlife contact). By setting reporting bias in to the maximum known for any wild mammal, it is possible to identify geographic regions with the largest estimated number of 'missing viruses' and proportion of 'missing zoonoses' (*6*). Prioritizing sampling of these species and regions would therefore provide the highest rate of zoonotic viral discovery for the GVP. *2) Targeting EID hotspots.* Analysis of all infectious diseases reported as emerging in humans since 1960 shows that, when corrected for reporting bias, their geographic origins are correlated with wildlife biodiversity and human population density (*1*). The spatial distribution of these correlated drivers provides a geographic map of EID hotspots which represent the most likely regions for future spillover events. This analysis has been revised significantly, with a new database, new analytical approach, new measures of reporting effort, new validation methods, and specifically targeting zoonotic EIDs (*27*). The results demonstrate that, after accounting for reporting effort, zoonotic EID risk is elevated in forested tropical regions experiencing land-use changes and where wildlife biodiversity (mammal species richness), and human population density, are high. Apart from the correlation with mammal species richness which is already accounted for in the GVP targeting, these results could be used to identify initial GVP sampling sites. Novel viruses identified in wildlife at these sites would have the highest relative propensity to emerge, other traits being equal. *3) Syndromic surveillance at wildlife-livestock-human interfaces.* Targeting GVP sampling of wildlife to sites with repeated reports of undiagnosed outbreaks in humans or livestock may increase the probability of identifying causative agents of these outbreaks. For example, there are substantial available data on undiagnosed case clusters and outbreaks of encephalitis and other discrete syndromes with potential infectious etiology from the literature and from internet-based reporting. During the late 1990s,

analyses of these reports for South Asia suggested that some outbreaks may be caused by Nipah virus – at the time not known in the region. This provided the rationale for a survey of fruit bats in Northern India (the reservoirs for Nipah virus elsewhere), which led to the first report of Nipah virus in bats in India (*34*). Syndromic surveillance allowed identification *post hoc*, of the causative agent of outbreaks in South Asia (*35, 36*). Similarly, syndromic surveillance of livestock has been used to identify diseases of significance to livestock production or public health (*37-40*). Data on undiagnosed clusters of potential infectious etiology in livestock could be used to target the planned livestock sampling in the GVP. These approaches may provide more direct benefits from the GVP for public health and food security. *4) Initial targeting of RNA viruses.* RNA viruses have caused 94% of the zoonoses documented from 1990 to 2010, 28 times higher than the proportion of RNA viruses detected among all vertebrate viruses (*30*). Therefore, to maximize the early detection of likely zoonotic viruses, GVP testing could target RNA viral families, including retroviruses. Alternatively, following testing for all 24 viral families, novel viruses could be triaged initially for further characterization based on their RNA vs. DNA genome. *5) Economies of scale.* Economies of scale would potentially reduce the cost of the GVP, providing that issues of access to sampling sites, sample movement and intellectual property are negotiated equitably. For example, modeling could be used to identify the maximum return (i.e. number of unique viruses identified from wildlife species, or number of unique wildlife species adequately sampled) on investment (number of sampling sites and cost) per country. In regions with discrete pockets of biodiversity, algorithms used to identify sites for protected areas could be modified to increase GVP site selection efficiency (*41-43*). At a national scale, the presence of biodiversity hotspots across country boundaries could lead to repetitive sampling over two adjacent countries if targeting is based on country-by-country analyses. A regional (multinational) approach may reduce costs significantly while giving the same species coverage, and would allow for decisions on logistical efficiencies to drive sample site choice. Testing of samples in laboratories that already have a regional remit may increase efficiency and reduce costs of laboratory capacity building. Similarly, regional bioinformatics platforms could be used that would not only increase the capacity for analysis of sequence data, but also provide training in bioinformatics for scientists across a region. *6) Technological innovation.* The decade-long timescale of the GVP may allow for technological innovation to significantly reduce testing and bioinformatics costs and therefore to increase efficiencies. As an extreme example, the cost of sequencing a human genome fell from $100,000,000 to around $10,000 in 10 years (2002-2012), i.e. 4 orders of magnitude, exceeding projected increased efficiency by Moore's Law (*44*). Technological advances continue in biased and unbiased viral discovery technology (*45*), and are likely to increase testing efficiency and reduce costs (*46*).

Potential direct benefits of the GVP to reducing risk of zoonotic viral emergence

The GVP's goal of identifying the bulk of currently-unknown viruses with zoonotic potential from wildlife reservoirs will have substantial potential for preventing and controlling zoonotic viral emergence in the future. However, the development of biomedical countermeasures (therapeutic drugs or vaccines) would require significant financial investment beyond that costed out for the GVP, and significant time following discovery (*47-49*). For example, substantial research on influenza viruses over the past

few decades has yielded a wealth of data on virology, immunology, pathology and epidemiology. This has led to strategies to predict the emergence of seasonal flu strains and to develop vaccines for them in advance of their emergence. However, their efficacy appears to be variable (*50*), and vaccine development to the recently emerging H1N1 pandemic strain lagged its global spread, albeit that speed of production was greatly enhanced (*51, 52*).

Despite the challenges of countermeasure development, there are a number of ways that the GVP could rapidly and feasibly improve efforts to pre-empt zoonotic viral emergence, more rapidly diagnose cases, and intervene earlier in outbreaks: *1) Greater diversity of viral reagents for countermeasure development.* Having the sequence data for thousands, rather than a few, viruses from a single family could enhance our understanding of the efficacy of novel therapeutics, vaccines or other countermeasures to a wider range of targets. For example, following the emergence of SARS-CoV, a large number of diverse related coronaviruses has been discovered from bats, including a group closely related to SARS - the bat SARS-like coronaviruses (SL-CoVs). The spike proteins of a number of recently-discovered SL-CoVs have ability to bind to the human receptor for SARS-CoV, and to infect and replicate SARS-like disease in mice when inserted into a mouse-adapted SARS-CoV backbone (*13, 53*). However, both a vaccine and monoclonal therapy that reduces SARS-CoV infection in this model failed to neutralize and protect from infection with the SL-CoV chimera *(54)*. Thus, expanding our knowledge of the diversity of related SL-CoVs under a GVP program may provide more effective testing protocols for SARS therapeutics and prophylactics; *2) Capacity building.* To facilitate the scaling up of current surveillance and viral discovery programs, the GVP would necessarily require training of One Health teams in regions most likely to experience novel outbreaks of emerging diseases. In the USAID EPT/PREDICT project, teams of trained wildlife biologists capable of safely sampling wildlife during outbreak conditions have assisted in outbreaks of yellow fever in Africa, Ebola virus disease in West Africa, Nipah virus in Bangladesh and a range of others (*55*). Scaling up of in-country One Health capacity would help more rapidly identify the reservoirs of known or novel viruses causing outbreaks in EID hotspots, and identify exclusion measures or other strategies to reduce risk of further spillover; *3) Enhancing rapid diagnosis during outbreaks.* The identification of relatives of known zoonotic viruses in new regions may enhance the speed and capacity to diagnose at-risk populations. For example, the identification of novel and distinct henipaviruses in bats in Africa has led to testing for, and evidence of, their transmission to pigs and humans in the region (*56-58*). This group of viruses can now be included in investigation of future outbreaks of Nipah- or Hendra-like disease in livestock in the region. *4) Designing risk mitigation policies.* The metadata collected by the GVP on viral-host ecology, viral distribution, seasonal dynamics, proximity to dense human populations and travel centers, and to communities with cultural practices that increase risk of spillover, can be used to enhance public health. Sequence data from new relatives of known agents can be used to increase the breadth of coverage of PCR-based diagnostic tests. Ecological and human behavioral data collected in the GVP will help refine our mechanistic understanding of spillover, and can be used to develop novel interventions. For example, the USAID EPT/PREDICT has identified sites in Yunnan Province, China where cave-dwelling bats harbor a high diversity of

novel SARS-like coronaviruses, including some capable of infecting human cells and causing illness in humanized mice (*13, 54, 59*). Ecological data collected during site evaluation prior to wildlife sampling, and human behavioral data collected routinely in the USAID EPT/PREDICT project identified human communities with significant exposure to bats from these caves. Targeting these communities for serological surveillance to SARS-like coronaviruses may help identify early evidence of spillover prior to full-scale emergence. Identification of potential zoonoses in hunted, traded and consumed wildlife species may provide a public health rationale for interventions (e.g. public education programs, enhanced biosecurity in farming systems and markets, policy interventions in wildlife trade) to reduce the risk of spillover. These mitigation strategies may also help prevent spillover of yet-to-be identified viruses within the same risk pathway or interface. The GVP metadata will also allow refinement of predictive models of zoonotic disease emergence, and viral host ranges (*1, 6, 11, 27*). For most of these, gaps in globally relevant datasets lower their predictive capacity, something that would be addressed by ground-truthing with much-expanded data on viral-host relationships and geographical distribution. *5) Strategies to triage potentially pandemic viruses for enhanced characterization and risk mitigation policies.* It would not be feasible to design and target control and intervention programs against all novel viruses identified by the GVP, and strategies to distinguish which novel viruses have true pandemic potential are rudimentary at present (*14*). Knowledge of influenza diversity and evolution has led to seasonal vaccine development based partly on predictive modeling of next year's dominant strain, but this is enhanced by a great deal of data on influenza viruses recently and currently circulating in people. To begin to address this issue, the USAID EPT/PREDICT project has developed an approach that characterizes the relative importance of newly-discovered viruses based on viral-dependent traits (e.g. proportion of known zoonoses per viral family, phylogenetic relatedness of known zoonoses, host breadth/plasticity,) and viral-independent traits (e.g. host species geographic range, abundance and proximity to people; viral prevalence in host; location of host in an EID hotspot) (*11*). Some of these approaches have been shown to predict zoonotic potential from large databases of virus-host relationships (*6, 30*). In the USAID EPT/PREDICT project, this strategy is used to identify which viruses are most likely to be potentially zoonotic, and to allocate resources for further in-depth characterization (*11, 55*). This approach was used successfully to further characterize the spike proteins of SARS-like coronaviruses discovered in this project, assess their capacity to bind to human cells, and conduct behavioral and serological surveys to assess the potential that spillover has occurred into the human population (*54, 59*). We propose that if this triage approach is used in the GVP, it will not only provide a preliminary framework for risk assessment, but will ultimately involve orders magnitude more data on viral sequence, geographic and host distribution, human behavioral and ecological risk and other parameters. This significant scaling up of available data will underpin a strategy to produce more realistic projections of zoonotic and pandemic potential.

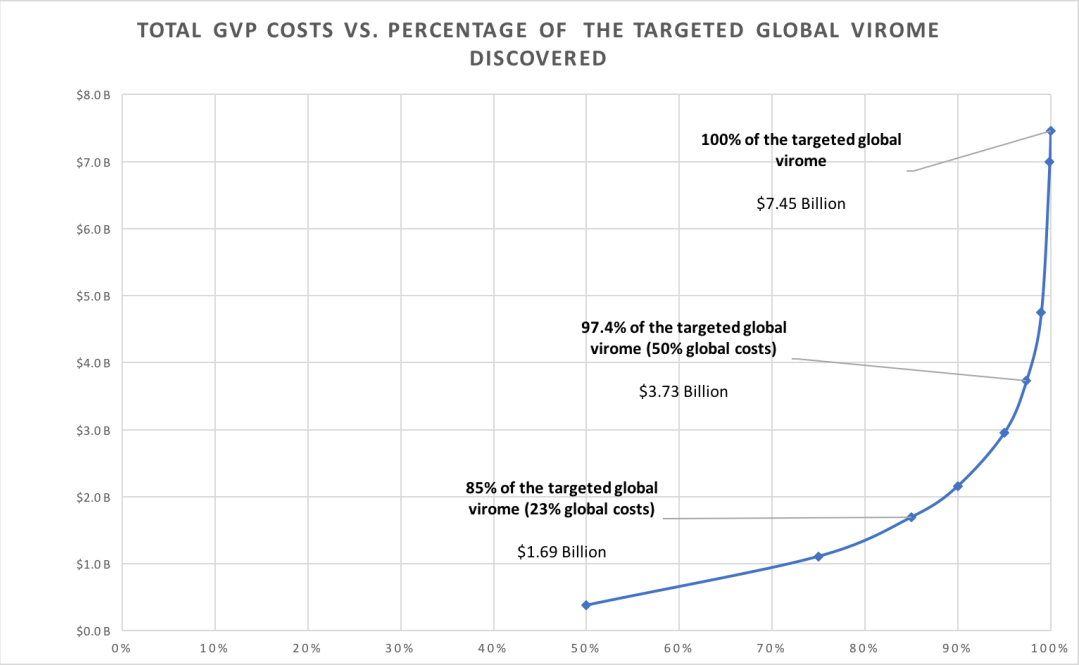

Fig. S1. Cost estimates of the GVP compared to the percentage of the targeted global virome discovered. The targeted global virome includes sampling from all 5,291 terrestrial mammal species and testing for 25 viral families of zoonotic potential, 871 species of water birds tested for influenza viruses, and limited (n=50) sampling of domestic animals and testing for 25 viral families.

1. K. E. Jones, N. G. Patel, M. A. Levy, A. Storeygard, D. Balk, J. L. Gittleman, P. Daszak, Global trends in emerging infectious diseases. *Nature* **451**, 990–993 (2008). doi:10.1038/nature06536 Medline

2. J. L. Geoghegan, A. M. Senior, F. Di Giallonardo, E. C. Holmes, Virological factors that increase the transmissibility of emerging human viruses. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 4170–4175 (2016). doi:10.1073/pnas.1521582113 Medline

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, Human viruses: Discovery and emergence. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **367**, 2864–2871 (2012). doi:10.1098/rstb.2011.0354 Medline

5. S. J. Anthony, J. H. Epstein, K. A. Murray, I. Navarrete-Macias, C. M. Zambrana-Torrelio, A. Solovyov, R. Ojeda-Flores, N. C. Arrigo, A. Islam, S. Ali Khan, P. Hosseini, T. L. Bogich, K. J. Olival, M. D. Sanchez-Leon, W. B. Karesh, T. Goldstein, S. P. Luby, S. S. Morse, J. A. K. Mazet, P. Daszak, W. I. Lipkin, A strategy to estimate unknown viral diversity in mammals. *mBio* **4**, e00598–e13 (2013). doi:10.1128/mBio.00598-13 Medline

6. K. J. Olival, P. R. Hosseini, C. Zambrana-Torrelio, N. Ross, T. L. Bogich, P. Daszak, Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650 (2017). doi:10.1038/nature22975 Medline

7. N. D. Wolfe, C. P. Dunavan, J. Diamond, Origins of major human infectious diseases. *Nature* **447**, 279–283 (2007). doi:10.1038/nature05775 Medline

8. J. Pike, T. Bogich, S. Elwood, D. C. Finnoff, P. Daszak, Economic optimization of a global strategy to address the pandemic threat. *Proc. Natl. Acad. Sci. U.S.A.* **111**, 18519–18523 (2014). doi:10.1073/pnas.1412661112 Medline

9. J. F. Drexler, V. M. Corman, M. A. Müller, G. D. Maganga, P. Vallo, T. Binger, F. Gloza-Rausch, V. M. Cottontail, A. Rasche, S. Yordanov, A. Seebens, M. Knörnschild, S. Oppong, Y. Adu Sarkodie, C. Pongombo, A. N. Lukashev, J. Schmidt-Chanasit, A. Stöcker, A. J. B. Carneiro, S. Erbar, A. Maisner, F. Fronhoffs, R. Buettner, E. K. V. Kalko, T. Kruppa, C. R. Franke, R. Kallies, E. R. N. Yandoko, G. Herrler, C. Reusken, A. Hassanin, D. H. Krüger, S. Matthee, R. G. Ulrich, E. M. Leroy, C. Drosten, Bats host major mammalian paramyxoviruses. *Nat. Commun.* **3**, 796 (2012). doi:10.1038/ncomms1796 Medline

10. data.predict.global

11. S. S. Morse, J. A. K. Mazet, M. Woolhouse, C. R. Parrish, D. Carroll, W. B. Karesh, C. Zambrana-Torrelio, W. I. Lipkin, P. Daszak, Prediction and prevention of the next pandemic zoonosis. *Lancet* **380**, 1956–1965 (2012). doi:10.1016/S0140-6736(12)61684-5 Medline

12. G. Grard, J. N. Fair, D. Lee, E. Slikas, I. Steffen, J.-J. Muyembe, T. Sittler, N. Veeraraghavan, J. G. Ruby, C. Wang, M. Makuwa, P. Mulembakani, R. B. Tesh,

J. Mazet, A. W. Rimoin, T. Taylor, B. S. Schneider, G. Simmons, E. Delwart, N. D. Wolfe, C. Y. Chiu, E. M. Leroy, A novel rhabdovirus associated with acute hemorrhagic fever in central Africa. *PLOS Pathog.* **8**, e1002924 (2012). doi:10.1371/journal.ppat.1002924 Medline

13. X.-Y. Ge, J.-L. Li, X.-L. Yang, A. A. Chmura, G. Zhu, J. H. Epstein, J. K. Mazet, B. Hu, W. Zhang, C. Peng, Y.-J. Zhang, C.-M. Luo, B. Tan, N. Wang, Y. Zhu, G. Crameri, S.-Y. Zhang, L.-F. Wang, P. Daszak, Z.-L. Shi, Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. *Nature* **503**, 535–538 (2013). doi:10.1038/nature12711 Medline

14. J. L. Geoghegan, E. C. Holmes, Predicting virus emergence amid evolutionary noise. *Open Biol.* **7**, 170189 (2017). doi:10.1098/rsob.170189 Medline

15. C. X. Li *et al*., *eLife* **4**, e05378 (2015).

16. S. J. Anthony, A. Islam, C. Johnson, I. Navarrete-Macias, E. Liang, K. Jain, P. L. Hitchens, X. Che, A. Soloyvov, A. L. Hicks, R. Ojeda-Flores, C. Zambrana-Torrelio, W. Ulrich, M. K. Rostal, A. Petrosov, J. Garcia, N. Haider, N. Wolfe, T. Goldstein, S. S. Morse, M. Rahman, J. H. Epstein, J. K. Mazet, P. Daszak, W. I. Lipkin, Non-random patterns in viral diversity. *Nat. Commun.* **6**, 8147 (2015). doi:10.1038/ncomms9147 Medline

17. *Convention on Biological Diversity* (1992); https://www.cbd.int/convention/).

18. P. M. Kinnunen, A. Palva, A. Vaheri, O. Vapalahti, Epidemiology and host spectrum of Borna disease virus infections. *J. Gen. Virol.* **94**, 247–262 (2013). doi:10.1099/vir.0.046961-0 Medline

19. J. A. Richt, R. Rott, Borna disease virus: A mystery as an emerging zoonotic pathogen. *Vet. J.* **161**, 24–40 (2001). doi:10.1053/tvjl.2000.0533 Medline

20. J. F. Drexler, A. Geipel, A. König, V. M. Corman, D. van Riel, L. M. Leijten, C. M. Bremer, A. Rasche, V. M. Cottontail, G. D. Maganga, M. Schlegel, M. A. Müller, A. Adam, S. M. Klose, A. J. Carneiro, A. Stöcker, C. R. Franke, F. Gloza-Rausch, J. Geyer, A. Annan, Y. Adu-Sarkodie, S. Oppong, T. Binger, P. Vallo, M. Tschapka, R. G. Ulrich, W. H. Gerlich, E. Leroy, T. Kuiken, D. Glebe, C. Drosten, Bats carry pathogenic hepadnaviruses antigenically related to hepatitis B virus and capable of infecting human hepatocytes. *Proc. Natl. Acad. Sci. U.S.A.* **110**, 16151–16156 (2013). doi:10.1073/pnas.1308049110 Medline

21. IUCN, "The IUCN Red List of Threatened Species. Version 2015.4.," (2015).

22. Birdlife International, "BirdLife Taxonomic Checklist v8.0" (Birdlife International, 2016).

23. J. F. He *et al*.; Chinese SARS Molecular Epidemiology Consortium, Molecular evolution of the SARS coronavirus during the course of the SARS epidemic in China. *Science* **303**, 1666–1669 (2004). doi:10.1126/science.1092002 Medline

24. R. Aldabe, L. Suárez-Amarán, C. Usai, G. González-Aseguinolaza, Animal models of chronic hepatitis delta virus infection host-virus immunologic interactions. *Pathogens* **4**, 46–65 (2015). doi:10.3390/pathogens4010046 Medline

25. F. A. Lempp, Y. Ni, S. Urban, Hepatitis delta virus: Insights into a peculiar pathogen and novel treatment options. *Nat. Rev. Gastroenterol. Hepatol.* **13**, 580–589 (2016). doi:10.1038/nrgastro.2016.126 Medline

26. J. Taylor, M. Pelchat, Origin of hepatitis delta virus. *Future Microbiol.* **5**, 393–402 (2010). doi:10.2217/fmb.10.15 Medline

27. T. Allen, K. A. Murray, C. Zambrana-Torrelio, S. S. Morse, C. Rondinini, M. Di Marco, N. Breit, K. J. Olival, P. Daszak, Global hotspots and correlates of emerging zoonotic diseases. *Nat. Commun.* **8**, 1124 (2017). doi:10.1038/s41467-017-00923-8 Medline

28. N. R. Faria, A. Rambaut, M. A. Suchard, G. Baele, T. Bedford, M. J. Ward, A. J. Tatem, J. D. Sousa, N. Arinaminpathy, J. Pépin, D. Posada, M. Peeters, O. G. Pybus, P. Lemey, HIV epidemiology. The early spread and epidemic ignition of HIV-1 in human populations. *Science* **346**, 56–61 (2014). doi:10.1126/science.1256739 Medline

29. N. D. Wolfe, W. M. Switzer, J. K. Carr, V. B. Bhullar, V. Shanmugam, U. Tamoufe, A. T. Prosser, J. N. Torimiro, A. Wright, E. Mpoudi-Ngole, F. E. McCutchan, D. L. Birx, T. M. Folks, D. S. Burke, W. Heneine, Naturally acquired simian retrovirus infections in central African hunters. *Lancet* **363**, 932–937 (2004). doi:10.1016/S0140-6736(04)15787-5 Medline

30. C. Kreuder Johnson, P. L. Hitchens, T. Smiley Evans, T. Goldstein, K. Thomas, A. Clements, D. O. Joly, N. D. Wolfe, P. Daszak, W. B. Karesh, J. K. Mazet, Spillover and pandemic properties of zoonotic viruses with high host plasticity. *Sci. Rep.* **5**, 14830 (2015). doi:10.1038/srep14830 Medline

31. A. D. Luis, D. T. S. Hayman, T. J. O'Shea, P. M. Cryan, A. T. Gilbert, J. R. C. Pulliam, J. N. Mills, M. E. Timonin, C. K. R. Willis, A. A. Cunningham, A. R. Fooks, C. E. Rupprecht, J. L. N. Wood, C. T. Webb, A comparison of bats and rodents as reservoirs of zoonotic viruses: Are bats special? *Proc. Biol. Sci.* **280**, 20122753 (2013). doi:10.1098/rspb.2012.2753 Medline

32. T. J. Davies, A. B. Pedersen, Phylogeny and geography predict pathogen community similarity in wild primates and humans. *Proc. Biol. Sci.* **275**, 1695–1701 (2008). doi:10.1098/rspb.2008.0284 Medline

33. B. Longdon, J. D. Hadfield, C. L. Webster, D. J. Obbard, F. M. Jiggins, Host phylogeny determines viral persistence and replication in novel hosts. *PLOS Pathog.* **7**, e1002260 (2011). doi:10.1371/journal.ppat.1002260 Medline

34. J. H. Epstein, V. Prakash, C. S. Smith, P. Daszak, A. B. McLaughlin, G. Meehan, H. E. Field, A. A. Cunningham, Henipavirus infection in fruit bats (*Pteropus giganteus*), India. *Emerg. Infect. Dis.* **14**, 1309–1311 (2008). doi:10.3201/eid1408.071492 Medline

35. T. L. Bogich, S. J. Anthony, J. D. Nichols, Surveillance theory applied to virus detection: A case for targeted discovery. *Future Virol.* **8**, 1201–1206 (2013). doi:10.2217/fvl.13.105

36. T. L. Bogich, S. Funk, T. R. Malcolm, N. Chhun, J. H. Epstein, A. A. Chmura, A. M. Kilpatrick, J. S. Brownstein, O. C. Hutchison, C. Doyle-Capitman, R. Deaville, S. S. Morse, A. A. Cunningham, P. Daszak, Using network theory to identify the causes of disease outbreaks of unknown origin. *J. R. Soc. Interface* **10**, 20120904 (2013). doi:10.1098/rsif.2012.0904 Medline

37. I. A. Bisson, B. J. Ssebide, P. P. Marra, Early detection of emerging zoonotic diseases with animal morbidity and mortality monitoring. *EcoHealth* **12**, 98–103 (2015). doi:10.1007/s10393-014-0988-x Medline

38. C. Dupuy, A. Bronner, E. Watson, L. Wuyckhuise-Sjouke, M. Reist, A. Fouillet, D. Calavas, P. Hendrikx, J.-B. Perrin, Inventory of veterinary syndromic surveillance initiatives in Europe (Triple-S project): Current situation and perspectives. *Prev. Vet. Med.* **111**, 220–229 (2013). doi:10.1016/j.prevetmed.2013.06.005 Medline

39. A. Elbers, R. Knutsson, Agroterrorism targeting livestock: A review with a focus on early detection systems. *Biosecur. Bioterror.* **11** (Suppl 1), S25–S35 (2013). doi:10.1089/bsp.2012.0068 Medline

40. T. Mwangi, K. Njenga, E. Otiang, P. Munyua, E. Ogola, T. Marsh, J. Yoder, S. Noh, T. McElwain, G. Palmer, Viewing human health through the lens of livestock health: Linked human and animal syndromic surveillance. *Int. J. Infect. Dis.* **21**, 36–36 (2014). doi:10.1016/j.ijid.2014.03.488

41. D. Newburn, S. Reed, P. Berck, A. Merenlender, Economics and land-use change in prioritizing private land conservation. *Conserv. Biol.* **19**, 1411–1420 (2005). doi:10.1111/j.1523-1739.2005.00199.x

42. S. Polasky, E. Nelson, E. Lonsdorf, P. Fackler, A. Starfield, Conserving species in a working landscape: Land use with biological and economic objectives. *Ecol. Appl.* **15**, 1387–1401 (2005). doi:10.1890/03-5423

43. K. A. Wilson, M. I. Westphal, H. P. Possingham, J. Elith, Sensitivity of conservation planning to different approaches to using predicted species distribution data. *Biol. Conserv.* **122**, 99–112 (2005). doi:10.1016/j.biocon.2004.07.004

44. E. C. Hayden, Technology: The $1,000 genome. *Nature* **507**, 294–295 (2014). doi:10.1038/507294a Medline

45. T. Briese, A. Kapoor, N. Mishra, K. Jain, A. Kumar, O. J. Jabado, W. I. Lipkin, Virome capture sequencing enables sensitive viral diagnosis and comprehensive virome analysis. *mBio* **6**, e01491–e15 (2015). doi:10.1128/mBio.01491-15 Medline

46. J. L. Mokili, F. Rohwer, B. E. Dutilh, Metagenomics and future perspectives in virus discovery. *Curr. Opin. Virol.* **2**, 63–77 (2012). doi:10.1016/j.coviro.2011.12.004 Medline

47. T. J. Liang, T. M. Block, B. J. McMahon, M. G. Ghany, S. Urban, J.-T. Guo, S. Locarnini, F. Zoulim, K.-M. Chang, A. S. Lok, Present and future therapies of hepatitis B: From discovery to cure. *Hepatology* **62**, 1893–1908 (2015). doi:10.1002/hep.28025 Medline

48. M. S. Cohen, K. E. Muessig, M. K. Smith, K. A. Powers, A. D. M. Kashuba, Antiviral agents and HIV prevention: Controversies, conflicts, and consensus. *AIDS* **26**, 1585–1598 (2012). doi:10.1097/QAD.0b013e3283543e83 Medline

49. M. J. Brooks, J. J. Sasadeusz, G. A. Tannock, Antiviral chemotherapeutic agents against respiratory viruses: Where are we now and what's in the pipeline? *Curr. Opin. Pulm. Med.* **10**, 197–203 (2004). doi:10.1097/00063198-200405000-00009 Medline

50. M. T. Osterholm, N. S. Kelley, A. Sommer, E. A. Belongia, Efficacy and effectiveness of influenza vaccines: A systematic review and meta-analysis. *Lancet Infect. Dis.* **12**, 36–44 (2012). doi:10.1016/S1473-3099(11)70295-X Medline

51. J. Oxford, A. Gilbert, R. L. Williams, in *Influenza Vaccines for the Future, Second Edition,* R. Rappuoli, G. DelGiudice, Eds. (2011), pp. 115-147.

52. A. Abelin, T. Colegate, S. Gardner, N. Hehme, A. Palache, Lessons from pandemic influenza A(H1N1): The research-based vaccine industry's perspective. *Vaccine* **29**, 1135–1138 (2011). doi:10.1016/j.vaccine.2010.11.042 Medline

53. V. D. Menachery, B. L. Yount Jr., A. C. Sims, K. Debbink, S. S. Agnihothram, L. E. Gralinski, R. L. Graham, T. Scobey, J. A. Plante, S. R. Royal, J. Swanstrom, T. P. Sheahan, R. J. Pickles, D. Corti, S. H. Randell, A. Lanzavecchia, W. A. Marasco, R. S. Baric, SARS-like WIV1-CoV poised for human emergence. *Proc. Natl. Acad. Sci. U.S.A.* **113**, 3048–3053 (2016). doi:10.1073/pnas.1517719113 Medline

54. V. D. Menachery, B. L. Yount Jr., K. Debbink, S. Agnihothram, L. E. Gralinski, J. A. Plante, R. L. Graham, T. Scobey, X.-Y. Ge, E. F. Donaldson, S. H. Randell, A. Lanzavecchia, W. A. Marasco, Z.-L. Shi, R. S. Baric, A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **21**, 1508–1513 (2015). doi:10.1038/nm.3985 Medline

55. PREDICT Consortium, "Reducing Pandemic Risk, Promoting Global Health.," (One Health Institute, University of California, Davis, 2014).

56. D. T. S. Hayman, R. Suu-Ire, A. C. Breed, J. A. McEachern, L. Wang, J. L. N. Wood, A. A. Cunningham, Evidence of henipavirus infection in West African fruit bats. *PLOS ONE* **3**, e2739 (2008). doi:10.1371/journal.pone.0002739 Medline

57. D. T. S. Hayman, L.-F. Wang, J. Barr, K. S. Baker, R. Suu-Ire, C. C. Broder, A. A. Cunningham, J. L. N. Wood, Antibodies to henipavirus or henipa-like viruses in domestic pigs in Ghana, West Africa. *PLOS ONE* **6**, e25256 (2011). doi:10.1371/journal.pone.0025256 Medline

58. A. J. Peel, D. R. Sargan, K. S. Baker, D. T. S. Hayman, J. A. Barr, G. Crameri, R. Suu-Ire, C. C. Broder, T. Lembo, L.-F. Wang, A. R. Fooks, S. J. Rossiter, J. L. N. Wood, A. A. Cunningham, Continent-wide panmixia of an African fruit bat facilitates transmission of potentially zoonotic viruses. *Nat. Commun.* **4**, 2770 (2013). doi:10.1038/ncomms3770 Medline

59. B. Hu, L.-P. Zeng, X.-L. Yang, X.-Y. Ge, W. Zhang, B. Li, J.-Z. Xie, X.-R. Shen, Y.-Z. Zhang, N. Wang, D.-S. Luo, X.-S. Zheng, M.-N. Wang, P. Daszak, L.-F.

Wang, J. Cui, Z.-L. Shi, Discovery of a rich gene pool of bat SARS-related coronaviruses provides new insights into the origin of SARS coronavirus. *PLOS Pathog.* **13**, e1006698 (2017). doi:10.1371/journal.ppat.1006698 Medline