

# Search strategy has influenced the discovery rate of human viruses

Ronald Rosenberg<sup>1</sup>, Michael A. Johansson, Ann M. Powers, and Barry R. Miller

Division of Vector-Borne Diseases, Centers for Disease Control and Prevention, Fort Collins, CO 80521

Edited\* by Barry J. Beaty, Colorado State University, Fort Collins, CO, and approved June 27, 2013 (received for review April 18, 2013)

**A widely held concern is that the pace of infectious disease emergence has been increasing. We have analyzed the rate of discovery of pathogenic viruses, the preeminent source of newly discovered causes of human disease, from 1897 through 2010. The rate was highest during 1950–1969, after which it moderated. This general picture masks two distinct trends: for arthropod-borne viruses, which comprised 39% of pathogenic viruses, the discovery rate peaked at three per year during 1960–1969, but subsequently fell nearly to zero by 1980; however, the rate of discovery of non-arboviruses remained stable at about two per year from 1950 through 2010. The period of highest arbovirus discovery coincided with a comprehensive program supported by The Rockefeller Foundation of isolating viruses from humans, animals, and arthropod vectors at field stations in Latin America, Africa, and India. The productivity of this strategy illustrates the importance of location, approach, long-term commitment, and sponsorship in the discovery of emerging pathogens.**

emerging diseases | zoonoses

Awareness of emerging human pathogens as imminent public health threats has grown rapidly over the last 20 y (1). Annual references to “emerging disease” in the medical literature have climbed from 11 in 1993 to 313 in 2012 (2); a similar search of popular print media uncovered two references in 1993 and 241 in 2012 (3). It is not clear, however, that the pace of emergence has also accelerated. What constitutes an emerging disease is often broadly interpreted to include not only those caused by newly discovered human pathogens but also by known pathogens whose transmission dynamics have changed, as when they develop drug resistance or invade previously nonendemic regions (4). A number of factors have been identified as contributing to an apparent rise in emerging diseases, including increased contact with wildlife in developing regions, ease of global travel and trade, and changing land use (5). By one estimate, 335 “emergent events” occurred during 1940–2004 (6). If the definition is limited—as it will be in this report—to only newly recognized causes of human disease, viruses now predominate. Of the six classes of agents causing disease (prions, viruses, bacteria, fungi, protozoa, and helminths), viruses comprised 67% of the 87 pathogens discovered during 1980–2005 (7), a period that saw the first descriptions of HIV, severe acute respiratory syndrome, hepatitis C, and Nipah viruses. Nearly 85% of these emergent viruses had single-stranded RNA (ssRNA) genomes (7), which are prone to uncorrected errors during replication (8), and the majority were zoonoses (7), pathogens capable of being transmitted to people from other vertebrates.

In an analysis of 188 viruses first found to cause human disease between 1900 and 2005 (9), the rate at which new viruses were identified in humans appeared to quicken between 1950 and 1970 but then to slow. To examine if changes in approach or methods might account for the change in pace, we compiled a list of the viruses found to cause human illness from 1897, when the first virus infecting vertebrates (foot-and-mouth disease virus) was discovered, through 2010 (Table S1). Selection criteria are described in *Methods*. Briefly, we used source reports for each virus (Table S1) to document not only the year it was first

incriminated as a cause of human disease, but also the year of discovery in a nonhuman animal (if that preceded when it was found in humans), the country of discovery, and the institutional home of the discoverer. The primary criteria for inclusion were that a virus be a proven pathogen of humans and that it be a taxonomically accepted species. In some cases the inclusion or exclusion of a virus was a matter of judgment because of ambiguous evidence for pathogenicity or unresolved taxonomic status, but analyses of lists made with either more- or less-stringent inclusion criteria did not change the nature of our results (*Methods* and Fig. S1). Of the 213 human viruses we designated, most also have known (60%) or presumed (8%) nonhuman vertebrate hosts, and 39% (83 of 213) are transmitted between humans or between animals and humans by arthropod vectors, such as mosquitoes or ticks (Table S1 and *SI Results and Discussion*).

## Results and Discussion

We plotted when each virus was first incriminated as a cause of human disease, as was done previously (9), but also when each was discovered (Fig. 1A). Of the 213 viruses, 31 (15%) were identified in nonhuman vertebrates and 27 (13%) in arthropod vectors before being incriminated as human pathogens (Table S1). The gap between the two cumulative trends (Fig. 1A) represents the lag in years (1–77 y) between discovery and incrimination (Table S1). Both trends are inflected more sharply upward in the early 1930s, when embryonated chicken eggs and suckling mice began to be used to culture unknown viruses, and in the early 1950s, when techniques for in vitro cell culture became widespread (Fig. 1A). Despite the further introduction of increasingly sophisticated diagnostic techniques, both trends began to slow in the 1970s. This slowing is especially evident in the data for discovery, to which we will limit further analysis. During 1960–1969, 47 viruses were discovered, compared with only 18 during 2000–2009. Of the 47 viruses, 68% (32 of 47) were arthropod-borne (“arboviruses”), a proportion much higher than 39% over the entire 114 y.

Both arbovirus and nonarbovirus discoveries accumulated at a similar pace (Fig. 1B) and rate (Fig. 1C) through the 1940s; by 1950 half (51%) the viruses now known to be pathogenic were arboviruses. The rapid growth in arbovirus discovery that subsequently occurred during the 1950s and 1960s was followed by a precipitous decline to nearly zero by 1980 (Fig. 1C). In contrast, nonarbovirus discovery remained consistent at approximately two new viruses yearly from the mid-1950s through 2010. Only 2% (2 of 83) of pathogenic arboviruses were discovered during the last 30 y of the study period (1981–2010), compared

Author contributions: R.R. designed research; R.R., A.M.P., and B.R.M. performed research; M.A.J. contributed new reagents/analytic tools; R.R. and M.A.J. analyzed data; and R.R., M.A.J., A.M.P., and B.R.M. wrote the paper.

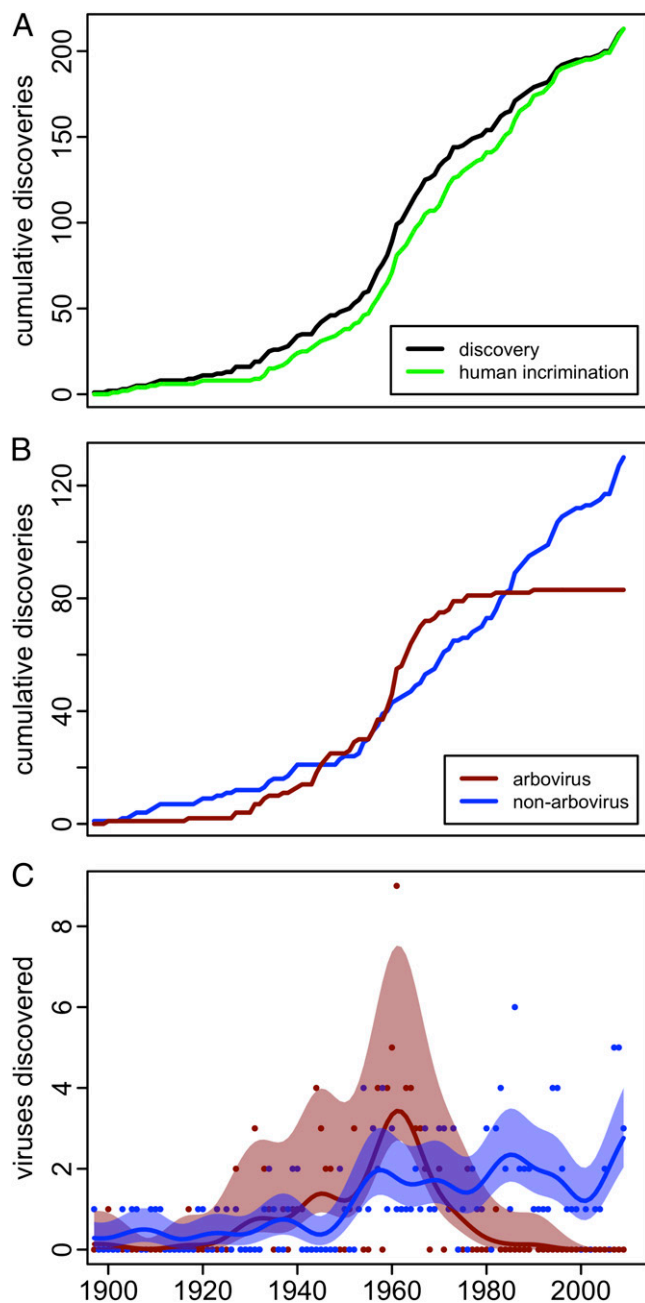
The authors declare no conflict of interest.

\*This Direct Submission article had a prearranged editor.

Freely available online through the PNAS open access option.

<sup>1</sup>To whom correspondence should be addressed. E-mail: viruslist@cdc.gov.

This article contains supporting information online at [www.pnas.org/lookup/suppl/doi:10.1073/pnas.1307243110/-DCSupplemental](http://www.pnas.org/lookup/suppl/doi:10.1073/pnas.1307243110/-DCSupplemental).



**Fig. 1.** The pace of human virus discovery 1897–2010. (A) The cumulative discoveries of human-pathogenic viruses in any organism (black) and their incrimination as a cause of human disease (green). (B) The cumulative discoveries of human-pathogenic arboviruses (red), and nonarboviruses (blue). (C) Yearly arboviruses and nonarboviruses discovered (red and blue points, respectively) with moving averages (lines) smoothed with a 10-y bandwidth Gaussian kernel and 95% confidence bands (shaded areas) determined by bootstrapping (10).

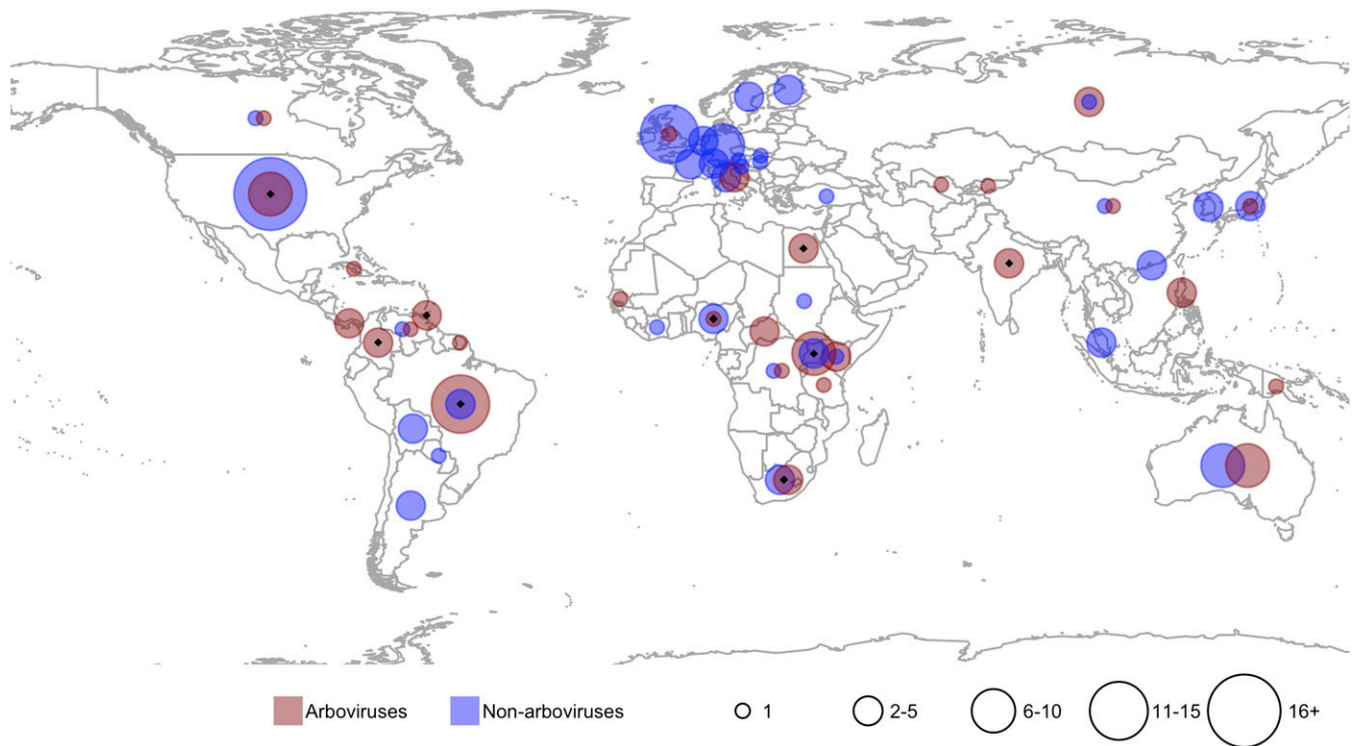
with 44% (57 of 130) of the nonarboviruses. The divergence of the two rates is statistically significant (Fig. 1C).

We can hypothesize two causes for the divergence in discovery rates between arboviruses and nonarboviruses: (i) by 1980 nearly all arboviruses pathogenic to humans had been found and new arboviral pathogens ceased to emerge, or (ii) there were differences in how arboviruses and nonarboviruses were discovered that no longer favored arbovirus discovery. Arboviruses appear to have at least as great an intrinsic capacity for emergence as do

nonarboviruses. Nearly all arboviruses (93%) have mutation-prone ssRNA genomes, compared with 57% of the nonarboviruses (Table S2). Most arboviruses (84%) are distributed among just three virus families (Table S2) but there is no compelling evidence that this reduces their capacity to evolve. The predominant family, the *Bunyaviridae*, which constitute 37% (31 of 83) of the arboviruses, have tripartite genomes especially susceptible to reassortment and potential changes in pathogenicity (11, 12). Recombination and mutation have also been documented in the other two major groups, *Flaviviridae* (13) and *Togaviridae* (14). All of the arboviruses are known or presumed zoonoses; even the few transmitted between humans by vectors, such as those causing dengue, have sylvatic transmission cycles (15). The passage through vectors and animals might favor relatively higher rates of synonymous change than in nonarboviruses (16), but there are a number of instances of single nucleotide polymorphisms contributing to arbovirus epidemics by enhancing infectivity to either the vector (17, 18) or animal host (19). More importantly, vectors often bridge gaps between humans and animals that would otherwise have no contact (20). The possibility that vector control has sufficiently reduced contact with humans to have stopped the introduction of novel arboviruses is belied by the failure of vector control practice to stem the worldwide resurgence of dengue, chikungunya, and other arboviruses (21).

Changes in the rate at which new species are discovered can result from changed approaches, such as investigation of novel habitats, development of new techniques, or redirection of resources (22). There was a striking difference in the geographic regions where arboviruses and nonarboviruses were first identified: 67% of arboviruses were discovered in Africa, Latin America/Caribbean, and India/Near East, whereas 68% of the nonarboviruses were discovered in North America and Europe (Fig. 2). There was also a distinct difference in the diversity of institutions responsible for making the discoveries (Table S3): 42% of the arboviruses were identified by staff of The Rockefeller Foundation (RF); in contrast, the largest proportion of nonarboviruses discovered by scientists employed by a single institution, the US National Institutes of Health, was 9%. Altogether, lead scientists working for 82 institutions were responsible for the discovery of the 130 nonarboviruses, compared with 25 institutions for the 83 arboviruses.

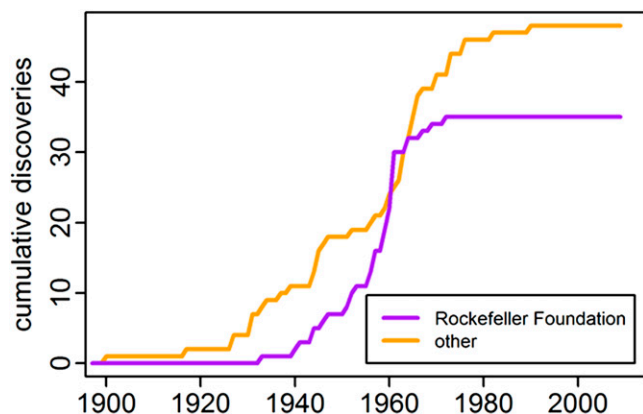
Beginning in 1916, the RF established field stations in Brazil, Nigeria, Colombia, and Uganda as part of a strategy to develop the means to eliminate yellow fever virus transmission, which led, by 1937, to the first yellow fever vaccine (23). During 1937–1950, RF scientists working at the yellow fever stations in Uganda, Colombia, and Brazil documented six previously unknown arboviruses—including West Nile virus—and beginning in 1951 RF instituted a 15-y program at laboratories in Egypt, Trinidad, South Africa, Brazil, Nigeria, Colombia, and India to find unrecognized arboviruses (23). The stations were staffed by RF and local professionals and equipped to hunt for and isolate viruses using state-of-the-art techniques, some of which had been pioneered by RF scientists. The operations, which were largely carried out in forested and rural areas, were modeled on the approach taken during the earlier yellow fever investigations and routinely attempted to isolate novel viruses from wild and sentinel animals, from wild-caught vectors, and from ill humans. This approach, which was also practiced by Institut Pasteur and others, was responsible for the high proportion of pathogenic arboviruses discovered first in animals: nearly half of human arboviruses (33% in arthropods, 13% in vertebrates) compared with 16% of nonarboviruses (Fig. S2). In particular, isolations from blood-feeding arthropods of viruses that displayed cytopathic effect in vertebrate tissue culture or were pathogenic to laboratory animals often preceded their incrimination as human pathogens. In addition to the 83 human arboviruses, more than



**Fig. 2.** Geographical distribution, by country, of discoveries of human-pathogenic arboviruses and nonarboviruses, 1897–2010. Black diamonds indicate countries with Rockefeller Foundation laboratories or field stations. Codiscoveries are represented in each country.

120 confirmed or probable arboviruses have been isolated from arthropods or vertebrates but not yet incriminated as infective to humans (24).

The sharp decline in discovery of human arboviruses was presaged by the predetermined end of the 15-y RF program in the mid-1960s (Fig. 3). We speculate there were two principal causes for the global stall in arbovirus discovery. First, the RF's influence on worldwide arbovirus research had become pre-eminent and its disappearance would have had a major impact on other institutions. The RF pioneered modern arbovirus research, provided short-term study grants, developed and disseminated new methods, arbitrated arbovirus classification, and provided a centralized type collection (25). The RF initiated the American Committee on Arthropod-Borne Viruses and the Catalogue of Arthropod-Borne Viruses of the World, iterations



**Fig. 3.** The cumulative discoveries of arboviruses by staff of The Rockefeller Foundation and by all other institutions.

of which are still active. RF investment during this period was estimated in 1973 as US \$30,000,000 (23). Second, the costly investment to describe novel human arboviruses in the tropics did not develop into a widely accepted post-World War public health imperative and no sponsor filled the vacuum the RF left (26). In the decade following the 1951 decision by the RF to support arbovirus discovery, the campaigns to eradicate malaria (1955) and small pox (1959) began, the first polio vaccine was licensed (1955), and the El Tor cholera pandemic started (1961). Although the RF made provision to move its collections and core staff to Yale University beginning in 1964, and to continue to provide for their support for a limited period, they did not further fund the field operations (25). Several of these continued active virology research programs but without support for the intensive, integrated approach characterized by the RF. No human arboviruses were isolated at any of the former RF supported field stations after 1970.

Our analysis was based on the well-documented discovery of viruses, but it highlights the general confounding influence of differing methodologies when judging historical trends or attempting to extrapolate to the future (9). The different patterns of human virus discovery were the consequence of fundamentally different investigative approaches. The nonarboviruses were predominately discovered in North America and Europe in humans, usually in the follow-up to disease outbreaks near major medical research institutions. In contrast, the arboviruses were predominately discovered in developing countries, often during integrated, geographically focused, longitudinal investigations. The influence that a single “big hitting” (27) sponsor had on the development and execution of the strategy to discover new arboviruses was profound, as were the consequences when its support ceased. The paucity of arbovirus discoveries from Southeast Asia, compared with those at similar latitudes in South America and Africa, might be linked to the absence of RF activity in the region.

Some arboviruses considered inconsequential at the time of their discovery have gained global importance as the conditions that favor emergence have increased. West Nile virus, discovered in Uganda in 1937 by the RF, rapidly spread across the United States after its introduction in 1999; in 2012 it produced severe neuroinvasive disease in about 3,000 Americans, which has been estimated to represent only about 1% of infections (28). The unprecedented 2004–2009 pandemic of chikungunya virus, also first described in East Africa, in 1957 at a laboratory established by the RF, infected at least 2 million people in the Indian Ocean region (29). Our analysis implies that not only has the emergence of arboviruses in the tropics been underestimated over the last 40 y but that the discovery of nonarboviruses has lagged as well because of the lack of systematic search. The development of faster, more sensitive methods for virus identification will continue to be important in discovering new pathogens, but strategy, support, and commitment for looking in the right places in the right ways will be critical to their most effective use.

## Methods

We initially compiled an inclusive list of 237 pathogenic viruses by consulting the most recent editions of major reference works (24, 30–32). With few exceptions, we examined the original published report of discovery for each virus and used the dates of publication as the dates of discovery (Table S1). Where the discoverers did not publish their results or longer than 5 y elapsed between discovery and publication, the earlier dates were used if known (Table S1). In some instances the discovery date was not when the virus was first isolated but when it was first determined to be a distinguishable species. For example, vaccinia and cowpox were frequently confused before 1939 (Table S1) and, although the cause of dengue fever was determined to be a virus in 1907 (33), it is not known which of the four viruses was investigated. We used the source documents to determine the institutional affiliation of the corresponding or senior author and the country of discovery.

- Lederberg J, Shope RE, Oaks SC (1992) *Emerging Infections: Microbial Threats to Health in the United States* (National Academies, Washington, DC).
- PubMed Health. Available at [www.ncbi.nlm.nih.gov/pubmed](http://www.ncbi.nlm.nih.gov/pubmed) (National Center for Biotechnology Information, Bethesda, MD). Accessed February 10, 2013.
- EBSCOhost. Available at <http://web.ebscohost.com> (EBSC, Ipswich, MA). Accessed February 12, 2013.
- McMichael AJ (2004) Environmental and social influences on emerging infectious diseases: past, present and future. *Philos Trans R Soc Lond B Biol Sci* 359(1447):1049–1058.
- Cleaveland S, Haydon DT, Taylor L (2007) Overviews of pathogen emergence: Which pathogens emerge, when and why? *Curr Top Microbiol Immunol* 315:85–111.
- Jones KE, et al. (2008) Global trends in emerging infectious diseases. *Nature* 451(7181):990–993.
- Woolhouse M, Gaunt E (2007) Ecological origins of novel human pathogens. *Crit Rev Microbiol* 33(4):231–242.
- Drake JW, Holland JJ (1999) Mutation rates among RNA viruses. *Proc Natl Acad Sci USA* 96(24):13910–13913.
- Woolhouse ME, et al. (2008) Temporal trends in the discovery of human viruses. *Proc R Soc B* 275(1647):2111–2115.
- Cowling A, Hall P, Phillips MJ (1996) Bootstrap confidence regions for the intensity of a Poisson point process. *J Am Stat Assoc* 91(436):1516–1524.
- Beatty BJ, Sundin DR, Chandler LJ, Bishop DH (1985) Evolution of bunyaviruses by genome reassortment in dually infected mosquitoes (*Aedes triseriatus*). *Science* 230(4725):548–550.
- Briese T, Bird B, Kapoor V, Nichol ST, Lipkin WI (2006) Batai and Ngari viruses: M segment reassortment and association with severe febrile disease outbreaks in East Africa. *J Virol* 80(11):5627–5630.
- Aaskov J, Buzacott K, Thu HM, Lowry K, Holmes EC (2006) Long-term transmission of defective RNA viruses in humans and *Aedes* mosquitoes. *Science* 311(5758):236–238.
- Weaver SC, et al. (1997) Recombinational history and molecular evolution of western equine encephalomyelitis complex alphaviruses. *J Virol* 71(1):613–623.
- Vasilakis N, Cardosa J, Hanley KA, Holmes EC, Weaver SC (2011) Fever from the forest: Prospects for the continued emergence of sylvatic dengue virus and its impact on public health. *Nat Rev Microbiol* 9(7):532–541.
- Woelck CH, Holmes EC (2002) Reduced positive selection in vector-borne RNA viruses. *Mol Biol Evol* 19(12):2333–2336.

Subsequently, we evaluated whether each virus on the initial list qualified as a distinct viral entity or species and the evidence that it caused human illness. For ~15% of the viruses on the initial list it was necessary to judge their qualifications for either or both criteria. Viruses were considered to be pathogens if they had been isolated from acutely ill people and shown to be the cause of the illness. Besides naturally acquired infections, including those only documented in immunocompromised people, those contracted during laboratory accidents were considered. Evidence for causality usually included studies demonstrating pathogenicity in humans or animal models and epidemiological or experimental evidence of transmission to humans. Because humans might be infected without overt disease, serological surveys indicating past exposure to a virus without evidence for pathogenicity were discounted, although rising convalescent antibody titers in individuals or seroconversions during well-documented epidemics were in some cases considered sufficient evidence. Putatively pathogenic viruses discovered during 2006–2010 were provisionally listed even if their role as pathogens had not yet been proved.

Systematic virus classification, especially at the species level, is frequently revised and the common professional use of virus names does not always match that accepted by the International Committee on Taxonomy of Viruses. Of the 213 listed viruses, 185 were listed as species in the ninth International Committee on Taxonomy of Viruses report (32). Of the remaining 28, 25 were classed as “tentative species,” “unassigned,” or “ungrouped,” and 3 (dengue-2, -3, -4) as subspecies (Table S1).

Table S1 is not meant to be authoritative but rather a means to investigate general trends in virus discovery. We tested how our judgment in compiling the list might affect the results by analyzing trends using the less discriminative list of 237 viruses and the limited list of 185 viruses (Fig. S1). There were no substantive differences between them.

**ACKNOWLEDGMENTS.** We thank the staff of the Public Health Library and Information Center of the Centers for Disease Control and Prevention for their help in finding and providing digital facsimiles of many references. Amy Lambert and Stuart Nichol of the Centers for Disease Control and Prevention provided taxonomic advice for some viruses.

- Brault AC, et al. (2004) Venezuelan equine encephalitis emergence: Enhanced vector infection from a single amino acid substitution in the envelope glycoprotein. *Proc Natl Acad Sci USA* 101(31):11344–11349.
- de Lamballerie X, et al. (2008) Chikungunya virus adapts to tiger mosquito via evolutionary convergence: A sign of things to come? *Virol J* 5:33.
- Brault AC, et al. (2007) A single positively selected West Nile viral mutation confers increased virogenesis in American crows. *Nat Genet* 39(9):1162–1166.
- Weaver SC, Barrett AD (2004) Transmission cycles, host range, evolution and emergence of arboviral disease. *Nat Rev Microbiol* 2(10):789–801.
- World Health Organization (2012) *Global Strategy for Dengue Prevention and Control, 2012–2020*. (World Health Organization, Geneva).
- Bebber DP, Marriott FH, Gaston KJ, Harris SA, Scotland RW (2007) Predicting unknown species numbers using discovery curves. *Proc R Soc B* 274(1618):1651–1658.
- Theiler M, Downs WG (1973) *Arthropod Borne Viruses of Vertebrates: An Account of the Rockefeller Foundation Virus Program, 1951–70* (Yale Univ Press, New Haven).
- Centers for Disease Control and Prevention. *International Catalog of Arboviruses Including Certain Other Viruses of Vertebrates*. Available at [www.cdc.gov/arbovat/index.asp](http://www.cdc.gov/arbovat/index.asp). Accessed April 1, 2013.
- Downs WG (1982) The Rockefeller Foundation virus program: 1951–1971 with update to 1981. *Annu Rev Med* 33:1–29.
- National Research Council (1987) *The US Capacity to Address the Tropical Infectious Disease Problem* (National Academy, Washington, DC).
- Bebber DP, et al. (2012) Big hitting collectors make massive and disproportionate contribution to the discovery of plant species. *Proc R Soc B* 279(1736):2269–2274.
- Centers for Disease Control and Prevention. *West Nile Virus Statistics, Surveillance, and Control*. Available at [www.cdc.gov/ncidod/dvbid/westnile](http://www.cdc.gov/ncidod/dvbid/westnile). Accessed April 1, 2013.
- Powers AM (2010) Chikungunya. *Clin Lab Med* 30(1):209–219.
- Knipe DM, Howley PM, eds (2006) *Fields Virology* (Lippincott Williams & Wilkins, Philadelphia, PA), 5th Ed.
- Mahy BWJ, van Regenmortel MHV, eds (2008) *Encyclopedia of Virology* (Elsevier, Oxford), 3rd Ed.
- King AMQ, Lefkowitz E, Adams MJ, Carstens EB, eds (2011) *Virus Taxonomy: Ninth Report of the International Committee on Taxonomy of Viruses* (Elsevier, Oxford).
- Ashburn PM, Craig CF (1907) Experimental investigations regarding the etiology of dengue fever. *J Infect Dis* 4:440–475.