

## PEARLS

# Advancing the central role of non-model biorepositories in predictive modeling of emerging pathogens

Jocelyn P. Colella<sup>1\*</sup>, Marlon E. Cobos<sup>1</sup>, Irene Salinas<sup>2,3</sup>, Joseph A. Cook<sup>2,4</sup>, The PICANTE Consortium<sup>†</sup>

**1** University of Kansas Biodiversity Institute and Department of Ecology & Evolutionary Biology, Lawrence, Kansas, United States of America, **2** University of New Mexico, Department of Biology, Albuquerque, New Mexico, United States of America, **3** Center for Evolutionary and Theoretical Immunology, University of New Mexico, Albuquerque, New Mexico, United States of America, **4** Museum of Southwestern Biology, University of New Mexico, Albuquerque, New Mexico, United States of America

<sup>†</sup> Membership of the PICANTE Consortium is provided in the Acknowledgments

\* [colella@ku.edu](mailto:colella@ku.edu)



The COVID-19 pandemic demonstrated the insufficiency of a reactive approach to emerging zoonotic pathogens. With spillover increasing in frequency as environments change and the human footprint continues to grow, pandemic prevention will require predictive models that can identify (i) potential zoonoses with a high likelihood of emergence and (ii) environmental or other features that may trigger a shift in host, vector, or pathogen baselines associated with emergence and/or spillover. Artificial intelligence (AI), and particularly its machine learning and deep learning branches, holds enormous potential for detecting shifts in large-scale biodiversity and disease datasets (genomic, ecological, geospatial, etc.) [1]. Such algorithms can be trained to identify subtle patterns in large volumes of data to yield insights into complex phenomena for which we have limited knowledge of the true cause (s) or predictor (s), as is the case for emerging infectious diseases.

Accuracy of AI or any other analytical approach, however, is limited by the quality, completeness, types, and biases of input data. Models are often hamstrung by unbalanced data. For example, if there are many host records, but few are infected by a pathogen, there may be insufficient information to identify a pattern. This complexity only increases when considering multiple host species, each of which has distinctive ecological requirements, generation time, life history, and dispersal ability (e.g., [2]). Using incomplete data to build predictive models can increase the number of candidate predictors, unnecessarily increasing computational complexity and processing time. Worse, it can produce overfit models that generalize poorly on new data and, in extreme cases, may increase false signal detection [3]. Thorough exploration of variable associations prior to inclusion in a larger model and transparency regarding the limitations of model extrapolation are critical (e.g., [4,5]).

Most emerging diseases in humans come from non-model organisms; therefore, understanding the ecology and evolution of those host species in the wild will be key to identifying features most informative for disease modeling and risk assessment. We focus on the enormous but relatively untapped potential of non-model host biorepositories and their associated databases to fuel predictive modeling of host–pathogen interactions, emergence risk, and pandemic potential across human–animal–environment interfaces. Non-model biorepositories, in this context, including natural history, agricultural, and other biodiversity collections that preserve biological materials (i.e., samples, specimens) in perpetuity. The metadata associated with the specimens warehoused in non-model biorepositories are served publicly online

## OPEN ACCESS

**Citation:** Colella JP, Cobos ME, Salinas I, Cook JA, The PICANTE Consortium (2023) Advancing the central role of non-model biorepositories in predictive modeling of emerging pathogens. *PLoS Pathog* 19(6): e1011410. <https://doi.org/10.1371/journal.ppat.1011410>

**Editor:** Neal Silverman, University of Massachusetts, Worcester, UNITED STATES

**Published:** June 15, 2023

**Copyright:** © 2023 Colella et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** PICANTE and this work were supported by the National Science Foundation Grant No. 2100955 (JPC, IS, JAC). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** The authors have declared that no competing interests exist.

through museum databases such as Arctos ([arctos.database.museum](http://arctos.database.museum)) and Symbiota ([symbiota.org](http://symbiota.org)), or data aggregators like the Global Biodiversity Information Facility (GBIF, [gbif.org](http://gbif.org)), iDigBio ([idigbio.org](http://idigbio.org)), and VertNet ([vertnet.org](http://vertnet.org)), some of which are interoperable with other digital data streams (GenBank, MorphoBank, etc.) derived from physical specimens. Through the newly established PICANTE (Pathogen Informatics Center: Analysis, Networking, Translation, and Education) initiative, we advocate for a multidisciplinary and, importantly, proactive approach to pandemic prediction and prevention (e.g., [6]) that emphasizes strategic expansion of non-model biorepositories, particularly in biodiverse countries, to stimulate and refine predictive modeling of emerging diseases.

## Going beyond serology

Serological tests or immunoassays screen blood for the presence of specific antibodies as an indication of historical infection. Such tests are a common tool used in disease surveillance because they are fast and affordable, but non-lethal serological investigations rarely, if ever, archive “voucher” specimens (e.g., preservation of host species and/or their tissues in a permanent biorepository) [7]. Thus, the disconnect between biorepositories and biomedical surveillance has created major gaps in knowledge and, critically, in biorepository sampling infrastructure that now limits our understanding of the ecology and evolution of emerging zoonoses and their wild hosts [8].

Serology requires commercially produced (“secondary”) antibody reagents capable of binding to specific immunoglobulins produced by the host. Yet, even within vertebrates, commercial reagents are often unavailable for non-model host species or are, potentially, less specific. As a result, critical model parameters such as baseline pathogen prevalence, transmission pathways, evidence of pathogenesis, rates of morbidity, occurrence of maternal immunity (i.e., placental, colostral transfer), and rates of recovery (e.g., [9]) remain unknown for most non-model wild hosts. In contrast, preservation of a holistic voucher specimen at the time of sampling would allow preliminary serological results to be revisited or extended at a later date, using more sensitive or updated technologies (e.g., whole-genome sequencing, metagenomics, RNAseq, environmental DNA) to fill information gaps. The date of serological screening is also not equivalent to the date of infection, as infection could have occurred months prior to screening. Thus, seropositivity cannot precisely associate a pathogen with the ecological or environmental conditions recorded at the time of sampling. Instead, models based on seroprevalences must consider the breadth of conditions present during the hosts’ life span, which severely limits resolution into biotic, abiotic, and anthropic factors that may contribute to outbreak or spillover. Again, if voucher tissues are preserved, reverse transcription polymerase chain reaction can be used to identify active infections among seropositive individuals, with sequencing then possible on preserved host tissues to identify strains, probe immune responses, and investigate pathogen evolution in detail.

## Work smarter, start with extension

The World Health Organization’s Blueprint [10] aims to “improve coordination between scientists and global health professionals to accelerate research and development related to emerging diseases” [11]. The Blueprint prioritizes pathogens based on their potential to cause the next global pandemic [12]. In this case, pandemic potential is measured based on information often available through non-model biorepositories, such as insights into human–animal interfaces, the evolutionary potential and geographic range of the pathogen, and existence of previous immunity [11]. The WHO Blueprint could be made more effective by including input from biorepository professionals with expertise in field collection methods, taxonomy

and systematics, and state-of-the-art long-term specimen preservation methods [7]. Recent, large-scale investigations have screened >75,000 mammals for viruses, but preserved few to no physical specimens (e.g., [13]), even in cases where hosts could not be identified to species. Such an approach precludes verification of host and pathogen taxonomy and limits future extension of initial work using new tools in the rapidly evolving fields of genomics, immunology, or isotopic chemistry [14] and serves to highlight the enormous potential for biorepositories to synergize with the public health and biomedical communities.

### Leveraging biorepositories for pathogen prediction

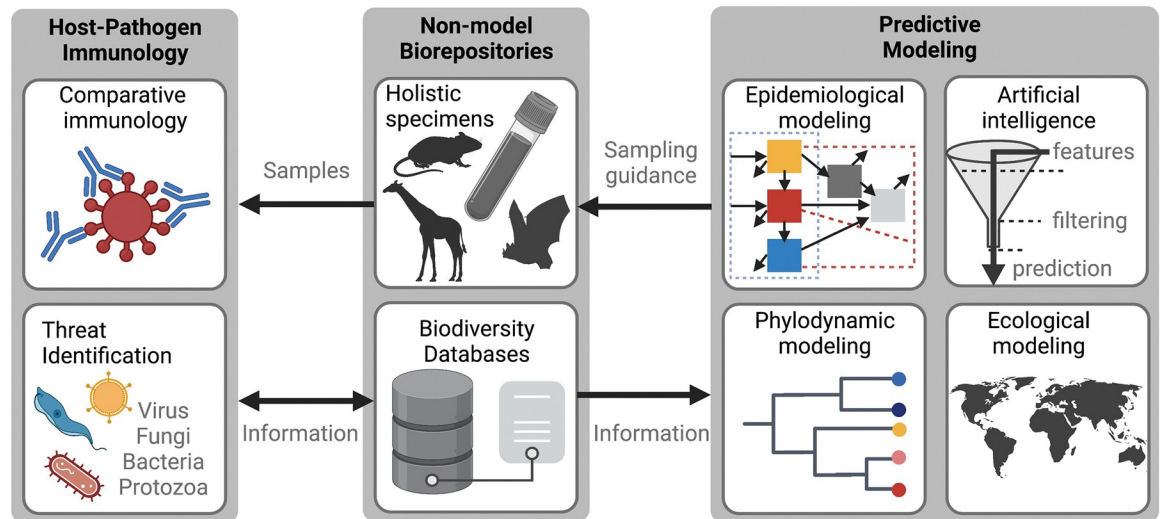
To maximize information gained, disease surveillance must be designed with verification, replication, and extension in mind [15]. Holistic collection and specimen vouchering [16,17] involve subsampling and archiving multiple parts of an organism to not only answer an initial question, but to also intentionally catalyze diverse scientific inquiry and facilitate integration across disciplines by tying newly derived information (e.g., genetic sequence data, serology results) back to a physical specimen record. Vouchering a subset or, ideally, all sampled taxa in a non-model biorepository ensures future verification, replication, and extension of research discoveries [15].

Biorepositories and their associated databases are rich, openly accessible sources of physical samples and digital data useful for diagnostic testing (e.g., serology, sequencing), monitoring change through time, and building predictive models of host–pathogen–environment interactions. Such collections are assembled by the scientific community over time, through the cumulative contributions of many researchers, agencies, and laboratories that sample natural systems to together produce a temporally deep, geographically broad, and taxonomically diverse global archive of biodiversity from which we can better understand host–pathogen–environment interactions. To be maximally useful, biodiversity databases associated with non-model biorepositories must be openly available online, machine readable, and standardized (e.g., DarwinCore) to allow data from different sources to be combined to increase sample sizes and, therefore, statistical power [18]. Specimens, including frozen tissue resources or cultures, must be available by loan for use in research and diagnostic testing, conditional on compliance with international regulations, including equitable sharing of benefits with international partners [19].

Most modeling applications require balanced input data, yet biodiversity data streams are rarely balanced with respect to taxonomy, sex, or geography and often do not comprehensively represent the entire range of environmental conditions in which we are trying to make predictions. Examples include larger-bodied animals and species of conservation concern that are generally underrepresented in collections or oversampling of highly accessible areas due to logistical constraints [20]. The gap between traditional disease surveillance efforts and sample archival with biorepositories has exacerbated those biases by targeting specific geographic areas or host species following an outbreak. Thus, when model focus is narrowed to a particular place, time, or species, available information may be drastically reduced (e.g., [21]). Moving forward, vouchering specimens as a regular part of disease surveillance will help fill data gaps, even biases, and build foundational infrastructure for biodiversity and disease-related informatics research.

### Non-model host–pathogen interactions, ecology, and immunology

A pathogen can only cause disease when it encounters a susceptible host in an environment conducive to infection. Thus, capturing and understanding disease dynamics in nature requires knowledge of all 3 vertices of the epidemiological triad [22]: pathogens, hosts (plus,



**Fig 1. Concept diagram showing an example of how the physical samples and associated data available through non-model biorepositories can fuel diverse aspects of host-pathogen modeling, immunology, and phylodynamics, among other applications relevant to public health.** Figure generated with BioRender ([biorender.com](https://biorender.com)).

<https://doi.org/10.1371/journal.ppat.1011410.g001>

vectors), and environments. Before applying models to predict pathogen emergence, it is best practice to first explore the scale (e.g., temporal, geographic, taxonomic), types of variables, and sampling intensity needed to reliably detect deviations from baseline conditions (e.g., [23,24]). Such information can guide strategic, holistic sampling to grow biorepository resources, most critically in biodiverse corners of the globe where pathogen emergence may be more likely [25]. This effort will require funding, expanded biodiversity infrastructure, expertise in holistic field sampling and non-model organism taxonomy, sample archiving with publicly accessible biorepositories, and collaborative, multidisciplinary perspectives (Fig 1). Ultimately, to be proactive, the entire pipeline, from sampling to analysis to policy action, must occur at a rate and scale relevant to public health. Top-down guidance that encourages researchers to contact non-model biorepositories early, as a key partner in disease surveillance, will be critical.

Once infected by a pathogen, transmission potential is determined by the host immune response. The diversity of immune systems on Earth is astounding, yet most immunology and, therefore, investigations of transmission potential have been based on relatively few model species. Pathogens of pandemic potential, however, are mostly hosted by non-model organisms, which have largely unexplored immune systems. The genomic era has brought unparalleled molecular knowledge of such non-model hosts, often illustrating substantial variation even among related species (e.g., [26]). As such, biorepository samples are invaluable resources that can illuminate variation in host susceptibility, transmission potential of wild species, and overall determine the pandemic potential of zoonotic pathogens. For example, the combination of genomes, transcriptomes, and newly developed cell lines [27–29] has propelled our understanding of how bats deal with viral pathogens and why they are common vectors of zoonotic diseases. Yet, this knowledge and equivalent tools are missing for most non-model hosts. Examples include raccoon dogs and Chinese bamboo rats, which have been recently implicated in the origins of the SARS-CoV-2 outbreak in Wuhan Markets [30]. Thus, predicting emergence and potential pandemics is inextricably linked to understanding immunity across a diversity of host species, and, if strategically built, biorepositories can be powerfully leveraged to fill current immunological knowledge gaps.

## Future directions

Pandemic prediction is both challenging and in its infancy. Successful prediction will require proactive multidisciplinary initiatives that intentionally contribute to and expand the sampling and informatic infrastructure of non-model biorepositories. Integrating new computational tools with biorepository archives can be powerfully used to (1) identify new pathogens with zoonotic potential, (2) understand key interactions among pathogens and their wild hosts, and (3) model host–pathogen interactions and risk landscapes in geographic, genomic, and environmental space. Transmission ultimately occurs at interfaces between people, animals, and environments and can be triggered by subtle shifts at any of those levels. We are now tasked to build the primary biodiversity infrastructure necessary to document and assess shifting interfaces and connect these data resources directly to computational pipelines to form an early warning system leading to community-level public health action and policies (DAMA protocol; [6]).

PICANTE is an initiative, centered at the University of New Mexico, which aims to change how the scientific community builds and uses biodiversity infrastructure to proactively identify and respond to zoonotic pathogen emergence. PICANTE is accomplishing this through a series of interdisciplinary collaborations that bridge biological science, engineering, computer science, and the social sciences towards pandemic prediction and prevention by developing affordable, rapid, and scalable screening methods; expanding biodiversity infrastructure and capacities in biodiverse countries; and building predictive models that leverage biodiversity, environmental, and human social and behavioral data to identify and then monitor high-risk interfaces through a global network of biorepositories. PICANTE is distinctive in its proactive approach to emerging diseases—i.e., identifying pathogens and shifts in baseline conditions prior to spillover—and, by design, the project workflow forms a positive feedback loop whereby additional sampling contributes directly to biorepositories, which, in turn, increases input data volumes and improves the accuracy of models derived from biodiversity data streams.

## Acknowledgments

We specifically recognize PICANTE's international partners, including the Instituto Conmemorativo Gorgas de Estudios de la Salud (Panamá), Centro de Investigación para la Salud en América Latina (CISEAL, Ecuador), Museo de Zoología de la Pontificia Universidad Católica del Ecuador (QCAZ, Ecuador), Instituto Nacional de Biodiversidad (INABIO, Ecuador), and the Center for Disease Control (CDC)–Panamá. We also thank the *Museums and Emerging Pathogens in the Americas* (MEPA) Project ECHO education and problem-solving network for fruitful discussion.

### **PICANTE Consortium members, alphabetical by Surname:**

Folashade Agosto, Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

Antonia Androski, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA

Akil Andrews, Department of Computer Science, University of New Mexico, Albuquerque, New Mexico, USA

Blas Armién, Department of Emergent and Zoonotic Diseases, Gorgas Memorial Institute for Health Studies, Panamá City, Panamá

Steven Bradfute, Center for Global Health and Department of Internal Medicine, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA

Santiago F. Burneo, Museo de Zoología, Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador, Quito, Ecuador



M. Alejandra Camacho, Museo de Zoología, Facultad de Ciencias Exactas y Naturales, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Judy L. Cannon, Molecular Genetics and Microbiology Department, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA

Jessica Carey-Webb, Spanish and Portuguese Cultural Studies, University of New Mexico, Albuquerque, New Mexico, USA

Juan Pablo Carrera-E, Instituto Nacional de Biodiversidad INABIO, Quito, Ecuador

Carlos Carrión Bonilla, Museo de Zoología, Escuela de Biología, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Lisette Chang, Center for Disease Control (CDC), Panamá

Marlon E. Cobos, Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

Joseph A. Cook, Museum of Southwestern Biology, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA

Jocelyn P. Colella, Biodiversity Institute and Department of Ecology and Evolutionary Biology, University of Kansas, Lawrence, Kansas, USA

Jaime A. Costales, Centro de Investigación para la Salud en América Latina, Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Darrell Dinwiddie, Department of Pediatrics, Division of Genetics, University of New Mexico, Health Sciences Center, Albuquerque, New Mexico, USA

Daryl Domman, Division of Infectious Diseases, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA

Jonathan L. Dunnum, Museum of Southwestern Biology, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA

Jeremy Edwards, Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico, USA

Samuel M. Goodfellow, Center for Global Health, Department of Internal Medicine, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA

Nicté Ordóñez Garza, Instituto Nacional de Biodiversidad INABIO, Quito, Ecuador

Emma Goldberg, Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos, NM; New Mexico Consortium, Los Alamos, New Mexico, USA

Publio González, Department of Emergent and Zoonotic Diseases, *Gorgas* Memorial Institute for Health Studies, Panamá City, Panamá

Mario Grijalva, Infectious and Tropical Disease Institute, Ohio University, Athens, Ohio, USA

Yi He, Chemistry and Chemical Biology, University of New Mexico, Albuquerque, New Mexico, USA

Eric P. Hoberg, Museum of Southwestern Biology, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA

Yan Lin, Department of Geography & Environmental Studies, Center for the Advancement of Spatial Informatics Research and Education (ASPIRE), University of New Mexico, Albuquerque, New Mexico, USA

Sofía Ocaña-Mayorga, Centro de Investigación para la Salud en América Latina (CISeAL), Pontificia Universidad Católica del Ecuador, Quito, Ecuador

Robert Miller, Center for Evolutionary and Theoretical Immunology, Department of Biology, University of New Mexico, Albuquerque, New Mexico, USA

Melanie Moses, Department of Computer Science, University of New Mexico, Albuquerque, New Mexico, USA

- Donald O. Natvig, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA
- Robert A. Nofchissey, Center for Global Health, Department of Internal Medicine, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA
- Adriana Romero-Olivares, Department of Biology, New Mexico State University, Las Cruces, New Mexico, USA
- Ethan Romero-Severson, Theoretical Biology & Biophysics, Los Alamos National Laboratory, Los Alamos, NM; New Mexico Consortium, Los Alamos, New Mexico, USA
- Paris Salazar-Hamm, Center for Global Health, Department of Internal Medicine, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA
- Irene Salinas, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA
- Sarigai Sarigai, Department of Geography & Environmental Studies, University of New Mexico, Albuquerque, New Mexico, USA
- Sarah Shrum Davis, Biology Department, University of New Mexico, Albuquerque, New Mexico, USA
- Nestor Sosa, Department of Internal Medicine, Division of Infectious Diseases, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA
- Andrew Skidmore, Department of Internal Medicine, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA
- Bruce Struminger, ECHO Institute and Department of Internal Medicine, Division of Infectious Diseases, Health Sciences Center, University of New Mexico, Albuquerque, New Mexico, USA
- Helen Wearing, Departments of Biology and Mathematics & Statistics, University of New Mexico, Albuquerque, New Mexico, USA
- Liping Yang, Department of Geography & Environmental Studies and Department of Computer Science, University of New Mexico, Albuquerque, New Mexico, USA

## References

1. Yang L, MacEachren AM, Mitra P. Geographical feature classification from text using (active) convolutional neural networks. In 2020 19th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE; 2020. p. 1182–1198.
2. Cecilia H, Vriens R, Wichgers Schreur PJ, de Wit MM, Métras R, Ezanno P, et al. Heterogeneity of Rift Valley fever virus transmission potential across livestock hosts, quantified through a model-based analysis of host viral load and vector infection. *PLoS Comput Biol*. 2022; 18(7):e1010314. <https://doi.org/10.1371/journal.pcbi.1010314> PMID: 35867712
3. Sagawa S, Raghunathan A, Koh PW, Liang P. An investigation of why overparameterization exacerbates spurious correlations. *International Conference on Machine Learning*. 2020. p. 8346–8356.
4. Alkische A, Cobos ME, Peterson AT, Samy AM. Recognizing sources of uncertainty in disease vector ecological niche models: an example with the tick *Rhipicephalus sanguineus sensu lato*. *Perspect Ecol Conserv*. 2020; 18:91–102.
5. Cobos ME, Peterson AT. Detecting signals of species' ecological niches in results and studies with defined sampling protocols: example application to pathogen niches. *Biodivers Inform*. 2022; 17:50–58.
6. Brooks DR, Hoberg EP, Boeger WA. *The Stockholm Paradigm: Climate Change and Emerging Disease*. University of Chicago Press; 2019. p. 423.
7. Thompson C, Phelps K, Allard M, Cook JA, Dunnum JL, Ferguson A, et al. Preserve a Voucher Specimen! The critical need for integrating natural history collections in infectious disease studies. *mBio*. 2021; 12:e02698–e02620. <https://doi.org/10.1128/mBio.02698-20> PMID: 33436435
8. Gilbert AT, Fooks AR, Hayman DTS, Horton DL, Müller T, Plowright R, et al. Deciphering serology to understand the ecology of infectious diseases in wildlife. *EcoHealth*. 2013; 10(3):298–313. <https://doi.org/10.1007/s10393-013-0856-0> PMID: 23918033
9. Keesing F, Ostfeld RS. Impacts of biodiversity and biodiversity loss on zoonotic diseases. *PNAS*. 2021; 118(17):e2023540118. <https://doi.org/10.1073/pnas.2023540118> PMID: 33820825

10. World Health Organization (WHO). 2022. Available from: <https://www.who.int/teams/blueprint> [cited December 2022].
11. Mehand MS, Al-Shorbaji F, Millett P, Murgue B. The WHO R&D Blueprint: 2018 review of emerging infectious diseases requiring urgent research and development efforts. *Antiviral Res.* 2018; 159:63–67.
12. Kieny MP, Salama P. WHO R&D Blueprint: a global coordination mechanism for R&D preparedness. *Lancet.* 2017; 389(10088):2469–2470.
13. Grange ZL, Goldstein T, Johnson CK, Anthony S, Gilardi K, Daszak P, et al. Ranking the risk of animal-to-human spillover for newly discovered viruses. *PNAS.* 2021; 118(15):e2002324118. <https://doi.org/10.1073/pnas.2002324118> PMID: 33822740
14. Dunnum JL, Yanagihara R, Johnson KM, Armien B, Batsaikhan N, Morgan L, et al. Biospecimen repositories and integrated databases as critical infrastructure for pathogen discovery and pathobiology research. *PLoS Negl Trop Dis.* 2017; 11:e0005133. <https://doi.org/10.1371/journal.pntd.0005133> PMID: 28125619
15. Colella JP, Stephens RB, Campbell ML, Kohli BA, Parsons DJ, Mclean BS. The open-specimen movement. *BioScience.* 2021a; 71(4):405–414.
16. Galbreath KE, Hoberg EP, Cook JA, Armien B, Bell KC, Campbell ML, et al. Building an integrated infrastructure for exploring biodiversity: field collections and archives of mammals and parasites. *J Mammal.* 2019; 100:382–393. <https://doi.org/10.1093/jmammal/gyz048> PMID: 31043762
17. Cook JA, Arai S, Armien B, Bates J, Carrión BC A, de Souza Cortez M. B, Soltis P S. Integrating Biodiversity Infrastructure into Pathogen Discovery and Mitigation of Emerging Infectious Diseases. *BioScience.* 2020; 70:531–534.
18. Wilkinson D, M Aalbersberg I J., Appleton G, Axton M, Baak A, Bouwman J. The FAIR guiding principles for scientific data management and stewardship. *Sci Data.* 2016; 3:160018. <https://doi.org/10.1038/sdata.2016.18> PMID: 26978244
19. Colella JP, Silvestri L, Suzán G, Weksler M, Cook JA, Lessa E. Engaging with the Nagoya Protocol on Access and Benefit-Sharing: recommendations for non-commercial biodiversity researchers. *J Mammal.* 2023.
20. Gularnick R, Van Cleve J. Strengths and weaknesses of museum and national survey data sets for predicting regional species richness: comparative and combined approaches. *Divers Distrib.* 2005; 11:349–359.
21. Xepapadeas A. The spatial dimension in environmental and resource economics. *Environ Dev Econ.* 2010; 15(6):747–758.
22. Last JM. *A Dictionary of Epidemiology.* Oxford University Press; 1988.
23. Glennon EE, Bruijning M, Lessler J, Miller IF, Rice BL, Thompson RN, et al. Challenges in modeling the emergence of novel pathogens. *Epidemics.* 2021; 37:100516. <https://doi.org/10.1016/j.epidem.2021.100516> PMID: 34775298
24. Salazar-Hamm PS, Montoya KN, Montoya L, Cook K, Liphardt S, Taylor JW, et al. Breathing can be dangerous: Opportunistic fungal pathogens and the diverse community of the small mammal lung mycobiome. *Front Fungal Biol.* 2022; 55.
25. Colella JP, Bates J, Burneo SF, Camacho MA, Carrión BC, Constable I, et al. Leveraging natural history biorepositories as a global, decentralized, pathogen surveillance network. *PLoS Pathog.* 2021b; 17(6): e1009583. <https://doi.org/10.1371/journal.ppat.1009583> PMID: 34081744
26. Moreno SD D Lama T M, Gutiérrez Guerrero Y. T Brown A M., Donat P, Zhao H, Dávalos LM. Large-scale genome sampling reveals unique immunity and metabolic adaptations in bats. *Mol Ecol.* 2021; 30(23):6449–6467. <https://doi.org/10.1111/mec.16027> PMID: 34146369
27. Zhou P, Tachedjian M, Wynne JW, Boyd V, Cui J, Smith I, et al. Contraction of the type I IFN locus and unusual constitutive expression of IFN- $\alpha$  in bats. *PNAS.* 2016; 113:2696–2701.
28. Kacprzyk J, Hughes GM, Palsson-McDermott EM, Quinn SR, Puechmaile SJ, O'Neill LAJ, et al. A potent anti-inflammatory response in bat macrophages may be linked to extended longevity and viral tolerance. *Acta Chiropterologica.* 2017; 19:219–228.
29. Brook CE, Boots M, Chandran K, Dobson AP, Drosten C, Graham AL, et al. Accelerated viral dynamics in bat cell lines, with implications for zoonotic emergence. *eLife.* 2020; 9:e48401. <https://doi.org/10.7554/eLife.48401> PMID: 32011232
30. Crits-Christoph A, Gangavarapu K, Pekar JE, Moshiri N, Singh R, Levy JI, et al. Genetic evidence of susceptible wildlife in SARS-CoV-2 positive samples at the Huanan Wholesale Seafood Market. Wuhan: Analysis and interpretation of data released by the Chinese Center for Disease Control. 2023. <https://doi.org/10.5281/zenodo.7754299>