Beyond Infection: Integrating Competence into Reservoir Host Prediction

Daniel J. Becker 1,*, Stephanie N. Seifert 2, and Colin J. Carlson 3

Most efforts to predict novel reservoirs of zoonotic pathogens use information about host exposure and infection rather than competence, defined as the ability to transmit pathogens. Better obtaining and integrating competence data into statistical models as covariates, as the response variable, and through postmodel validation should improve predictive research.

Predicting Reservoir Hosts

Epidemics of zoonotic disease beget questions about origins: from where did a pathogen originate? Rapid research during the coronavirus disease 2019 (COVID-19) pandemic has focused on identifying likely reservoir hosts (see Glossary) of severe acute respiratory syndrome coronavirus (SARS-CoV)-2. High nucleotide identity between this virus and related viruses found in horseshoe bats (e.g., Rhinolophus affinis) implies a wildlife origin, yet divergence of these viral lineages decades ago suggests hosts involved in spillover remain unknown [1]. Searching for closely related pathogens in wildlife is a labor-intensive process made all the more difficult by the diversity of possible hosts to sample. Given such challenges, ecologists and data scientists have developed statistical models to predict likely reservoir hosts, which can forecast sources of cross-species transmission and prioritize surveillance targets [2].

Predictive studies typically use parametric or semiparametric models (e.g., generalized linear or additive models) or machine learning algorithms (e.g., boosted regression trees), where the predictor matrix (i.e., X) includes life history, taxonomic, and geographic traits of hosts. In most cases, the response (i.e., y) is pathogen positivity or richness using data on the detection of pathogen antigen or pathogen-specific antibody [3]. Given many logistical hurdles in sampling wildlife, serology and PCR are often the first data available for predictive models. By indicating recent exposure, serology in particular remains useful for pathogens with short infectious periods and high temporal variability in infection [4]. However, such data do not necessarily reflect host competence, the ability to transmit pathogens [5]. In turn, even the best models using such data could generate spurious predictions. For example, broad serological sampling of bats has led to many bat–flavivirus associations in model-ready datasets, despite evidence that many bats are not competent reservoirs [6]. We here suggest that the use of serology and PCR data must be considered as a starting point for identifying likely reservoirs and emphasize how host competence data can be better integrated into all aspects of predictive research, including as covariates in the predictor matrix, as the response, and through post-model validation (Figure 1).

Host Competence Data

Competence can be defined most strictly as the host capacity to transmit pathogens to new susceptible hosts or arthropod vectors [7]. This definition restricts competence to only those infection processes that occur within the host: establishment of infection given exposure, pathogen development (i.e., replication), and pathogen survival until transmission. The former process is equivalent to host susceptibility, whereas the latter processes are contained within suitability and are mediated by resistance and tolerance mechanisms [5]. As competence encompasses several stepwise processes, this trait is continuous [8]. For some purposes, it may be approximated as binary [9].

As a composite trait, competence is difficult to determine in natural systems. Frequent detection of pathogen-specific antibodies or antigen alone cannot be interpreted as competence, as such patterns can instead reflect transmission from a true maintenance host. In wild hosts, competence can instead be inferred by detecting shedding of live virus (Figure 1). For vector-borne diseases, shedding may be less informative for ability to transmit. Instead, the susceptible state of vectors prior to feeding can be assumed in some systems, such that the infection status of engorged vectors can inform host competence. For example, as ticks are born free of Borrelia burgdorferi (the cause of Lyme borreliosis) and only become infectious from feeding, pathogen presence in engorged larvae implies competent hosts [8,9]. Competence can also be inferred by combining pathogen diagnostics of hosts and fed vectors with blood meal analyses [6].

In many cases, however, quantifying competence in field systems remains challenging. For example, despite bats being suggested reservoirs for Ebola virus and African Henipaviruses, live virus has yet to

Glossary

| Term | Definition |
|------|------------|
| Competence | Host ability to transmit pathogens to new susceptible hosts or arthropod vectors. |
| Reservoir host | Populations or species in which a pathogen can be maintained and that serve as a source of infection for the recipient host. |
| Serology | Detection of pathogen-specific antibodies in serum or plasma to infer exposure history. |
| Shedding | Release of pathogen in excreta (e.g., saliva, feces, and urine) following successful replication within the host. |
| Suitability | Ability of a host to support pathogen development (i.e., replication) and survival until transmission. |
| Susceptibility | Probability of an infection establishing in the host given exposure to a pathogen. |

Figure 1. Phylogenetic trees of zoonotic pathogens (Other viruses) and shed fingerprints (Bats) using data on the detection of pathogen antigen or pathogen-specific antibody [3].
be isolated despite frequent detection of antigen and antibody. Data on within-host components of competence, such as susceptibility and suitability (e.g., pathogen replication), could provide mechanistic covariates for predictive models using serology and PCR data as the response (Figure 1). As one example, species means in immunological covariates, such as leukocyte concentrations or ex vivo responses to pathogen challenge [10], could represent these within-host processes better than many currently used traits (e.g., body mass and fecundity). Host genetic variation can also shape susceptibility and suitability, especially for factors that enable pathogen entry into host cells and replication. For example, in silico methods can characterize receptor binding (i.e., informing susceptibility) and reveal host factors required for viral replication, which can be used as covariates or validation for models using PCR or serology responses. Lastly, in vivo experimental challenge studies can confirm pathogen replication and transmission to susceptible hosts or vectors, and results can serve as either the response or model validation.

in vitro studies are necessary to functionally validate in silico results and can further characterize host factors required for pathogen replication. As another example, reverse genetic assays were recently used to show functional compatibility of the spike glycoprotein from various novel SARS-like CoVs to enter cells with ACE2 from potential host species [12]. Lastly, in vivo pathogen challenge studies are the gold standard to determine shedding (i.e., productive infection) and onward transmission in a given host–pathogen system (Figure 1). For example, in the West Nile virus system, exposure of 25 bird species to infectious mosquitoes or virus challenge established greater competence of songbirds and shorebirds [13].

Integration into Predictive Research

These competence data can be integrated into different aspects of predictive research (Figure 1). Field-based measures of competence (i.e., detection of live virus or pathogens in recently fed vectors) can be prone to false negatives owing to short or highly variable periods of active infection or host seeking [4]. Longitudinal sampling is important to capture host shedding or transmission, yet such studies are logistically difficult. One clear contribution of predictive modeling based on serology and PCR is therefore to narrow the scope of hosts to then sample longitudinally for virus isolation and to understand transmission dynamics [2,3]. Resulting data on host competence could then allow confirming or refuting model predictions (Figure 1).

Similarly, within-host data can inform parameterizing (e.g., receptor use for susceptibility covariates) and validating predictive models. As in vivo pathogen challenge trials are costly, labor intensive, and can be limited to high-containment laboratories, they may be most useful for robustly testing

---

**Figure 1. Integrating Competence into Reservoir Host Predictive Models.** We take a simplified statistical model (in matrix notation, where $\beta$ represents regression coefficients and $\epsilon$ represents errors) and illustrate how data can be used as the response (arrows towards $y$), covariates (arrows towards $X$), and/or validation (arrows from $y$). Presence of pathogen antibody or antigen is commonly used as the response but conflates competence with exposure. Predictions from these models can be validated by field measures of competence, such as isolating live virus or diagnostics of certain arthropod vectors, both of which can also be more informative response variables. Both in silico and in vitro analyses can characterize receptor binding (i.e., informing susceptibility) and reveal host factors required for viral replication, which can be used as covariates or validation for models using PCR or serology responses. Lastly, in vivo experimental challenge studies can confirm pathogen replication and transmission to susceptible hosts or vectors, and results can serve as either the response or model validation.
model predictions (Figure 1). In one example, experimentally infected Egyptian fruit bats (*Roussettus aegyptiacus*) did not support *in vivo* replication of Nipah virus [14], despite prior serology-based predictive models suggesting this species could be a likely reservoir [3].

Iteratively integrating field studies, predictive models, and experimental trials could build a systematic dataset of host competence for select pathogens, which would facilitate including competence as the response in statistical models (Figure 1). To date, competence has been analyzed in several comparative frameworks (e.g., how host life history shapes the proportion of infected larval ticks per species [8]). Such data have generally yet to be included in larger predictive models, although doing so may generate more meaningful forecasts of likely reservoirs. In one recent example, the presence of infected larval ticks was integrated into machine learning algorithms to determine the taxonomic and life history correlates of competence for *Borrelia burgdorferi* across bird species, which allowed predicting likely but unsampled avian reservoirs [9]. A key research priority moving forward will be to determine how statistical models trained on serology, PCR, and competence response variables differ in performance and predictions.

**Improving Reservoir Host Prediction**

Efforts to predict reservoirs of zoonotic pathogens are expected to increase dramatically following the emergence of SARS-CoV-2 and the COVID-19 pandemic. On their own, traditional modeling approaches using data on antigen detection and seropositivity will likely provide limited insights into the underlying competence of proposed reservoir hosts. Being agnostic to host susceptibility and suitability, such models could fail to identify true reservoirs and lead to unnecessary responses (e.g., culling). Here, we have outlined how competence data could be better integrated into predictive research, through an iterative combination of testing model predictions, including within-host data as covariates, and eventually modeling competence itself as the response (Figure 1). Such work could produce more relevant forecasts of reservoirs, identify host traits with causative physiological links to infection, and enhance our understanding of how different wildlife species contribute to pathogen transmission in host communities. To accomplish such aims, we emphasize a broader need for close collaboration between ecologists, data scientists, microbiologists, and immunologists to ensure that data and methods are used and interpreted accurately and to foster better dialogue between prediction and validation.

**Acknowledgments**

D.J.B. was supported by an appointment to the Intelligence Community Postdoctoral Research Fellowship Program at Indiana University, administered by Oak Ridge Institute for Science and Education through an interagency agreement between the US Department of Energy and the Office of the Director of National Intelligence. S.N.S. was supported by the Intramural Research program of the National Institute of Allergy and Infectious Diseases. The authors were supported by funding to the Viral Emergence Research Initiative (VENERA) consortium, including NSF BiL 2021909 and a grant from Institut de Valorisation de Données (IVADO). We thank members of the Ketterson Laboratory and two anonymous reviewers for helpful feedback on earlier versions of this manuscript.

1. Department of Biology, Indiana University, Bloomington, IN, USA
2. Laboratory of Virology, Division of Intramural Research, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, MT, USA
3. Center for Global Health Science and Security, Georgetown University Medical Center, Washington, D.C., USA
4. Correspondence: danbeck@iu.edu (D.J. Becker).

References

1. Boni, M.F. et al. (2020) Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat. Microbiol. Published online July 28, 2020. https://doi.org/10.1038/s41564-020-0771-4
2. Becker, D.J. et al. (2019) The problem of scale in the prediction and management of pathogen spillover. Philos. Trans. R. Soc. B Biol. Sci. 374, 20190224
3. Piotrowit, R.K. et al. (2019) Prioritizing surveillance of Nipah virus in India. PLoS Negl. Trop. Dis. 13, e0007393
4. Piotrowit, R.K. et al. (2019) Sampling to elucidate the dynamics of infections in reservoir hosts. Philos. Trans. R. Soc. B Biol. Sci. 374, 20190336
5. Downs, C.J. et al. (2019) Scaling of host competence. Trends Parasitol. 35, 182–192
6. Vicente-Santos, A. et al. (2017) Neotropical bats that co-habit with humans function as dead-end hosts for dengue virus. PLoS Negl. Trop. Dis. 11, e0005507
7. Merrill, T.E.S. and Johnson, P.T. (2020) Towards a mechanistic understanding of competence: a missing link in diversity–disease research. Parasitology 147, 1159–1170
8. Ostfeld, R.S. et al. (2014) Life history and demographic drivers of reservoir competence for three tick-borne zoonotic pathogens. PLoS One 9, e107387
9. Becker, D.J. and Han, B.A. (2020) The macroecology and evolution of avian competence for *Borrelia burgdorferi*. bioRxiv Published online April 16, 2020. https://doi.org/10.1101/2020.04.15.040352
10. Downs, C.J. et al. (2020) The effects of body mass on immune cell concentrations of mammals. Am. Nat. 195, 107–114
11. Damas, J. et al. (2020) Broad host range of SARS-CoV-2 predicted by comparative and structural analysis of ACE2 in vertebrates. Proc. Natl. Acad. Sci. U. S. A. Published online August 21, 2020. https://doi.org/10.1073/pnas.201461117
12. Letko, M. et al. (2020) Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat. Microbiol. 5, 562–569
13. Komar, N. et al. (2003) Experimental infection of North American birds with the New York 1999 strain of West Nile virus. Emerg. Infect. Dis. 9, 311
14. Seiffert, S.N. et al. (2003) *Roussettus aegyptiacus* bats do not support productive Nipah virus replication. J. Infect. Dis. 221, S407–S413