

Improving Viral Sampling and Discovery: Viral Accumulation Curves

The USAID PREDICT project conducts viral surveillance in wildlife, domestic animals, and humans at a global scale. PREDICT teams in >30 countries worldwide have collected and tested samples from wildlife species and have discovered over 1,000 viruses from viral families known to threaten human health. Wildlife species are prioritized for surveillance based on existing scientific knowledge about the likelihood of viral spillover from these hosts to humans [1]. PREDICT aims to characterize viral diversity in mammals, which may represent hundreds of thousands of unknown viral species [2], before they emerge in people or domestic animals.

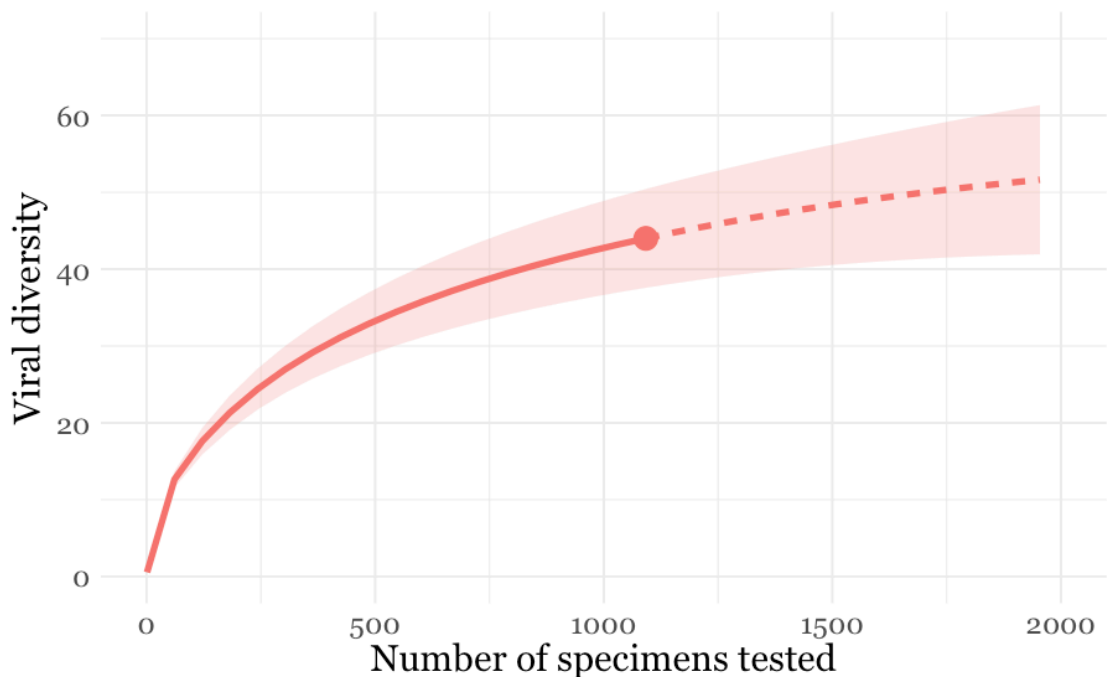
CHALLENGE: HOW MANY SAMPLES DO WE NEED TO COLLECT AND TEST TO FIND MOST OF THE VIRUSES IN A GIVEN HOST?

PREDICT seeks to discover currently unknown viruses in wildlife species that have been poorly sampled prior to this project. We need tools to tell us when we have discovered a majority of the viruses naturally circulating in a species so that we can stop sampling that species and move on to others.

SOLUTION: VIRAL ACCUMULATION CURVES

The PREDICT team uses statistical modeling techniques to solve this problem. Specifically, we are applying *species richness estimation methods* to determine when a wildlife species has been comprehensively sampled [3]. These estimates, typically used in ecological studies and modified here for viral discovery, can be expressed visually as *viral accumulation curves*.

Pteropus giganteus Viral Sampling (Nine Viral Families)



April 16, 2018

For details on methods or analysis contact:
PREDICTmodeling@ecohealthalliance.org



USAID
 FROM THE AMERICAN PEOPLE



Figure 1: Viral accumulation curve for the bat species *Pteropus giganteus*, adapted from Anthony et al. [4]. The x-axis of the graph shows the number of specimens, or samples, tested for nine viral families, and the y-axis indicates the number of unique viruses discovered in those samples. The solid line represents observed data, and the dashed portion represents statistically-derived estimates of viral diversity if sampling of *P. giganteus* is continued. The upward trend of the dashed line suggests that *P. giganteus* likely harbors more than the 44 viruses that were observed in this particular viral sampling effort.

The theory underlying these methods is that if sampling efforts repeatedly result in detection of a similar set of viruses, then sampling is likely to be near completion (i.e., most of the species present have been observed). In contrast, if sampling continues to generate observations of new viruses, then the pathogen community likely has many species yet to be observed. Early work by the PREDICT project used these techniques to estimate that the bat species *Pteropus giganteus* hosts a total of 58 viruses across nine viral families (Fig. 1) [4] and that rhesus macaques (*Macaca mulatta*) host 283 viruses across 14 viral families [5].

These pilot studies provide a proof-of-concept that has now been applied to other PREDICT data (Fig. 2). In addition to their usefulness in helping guide sampling strategies, comparison among accumulation curves can identify host species that consistently carry high viral diversity, potentially making them more important viral reservoirs. Viral observations can also be aggregated at coarser levels of host taxonomy to allow comparison of host genera or families, for example.

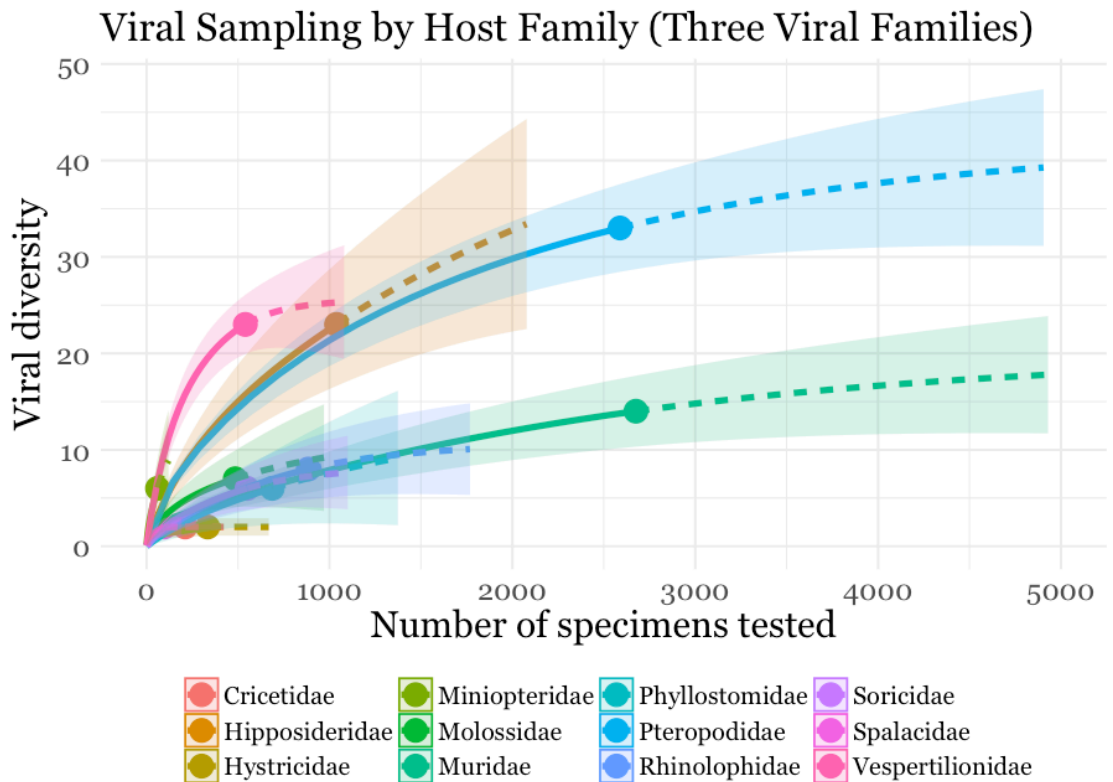


Figure 2: Viral accumulation curves using PREDICT data estimating viral species richness across host *families* for three major viral families: Coronaviruses, Influenzas, and Paramyxoviruses. These results suggest that, for these three viral families, some host taxa still have considerable viral diversity yet to be discovered (e.g., bats in the Hipposideridae which have an accumulation curve still trending upwards), whereas for others there is little unobserved viral diversity remaining.

CONCLUSIONS

- Species accumulation curves are a statistical method used to estimate the amount of unobserved diversity in a biological sample. For PREDICT, we have applied these techniques to our viral discovery data obtained from wildlife hosts.
- Application of viral accumulation curves using data from the PREDICT project, which is collected and tested in a consistent manner across species and countries, can improve our knowledge of which host taxa carry particularly rich viral assemblages and which host taxa require additional sampling efforts.
- These analytical tools support the overarching goals of PREDICT: to standardize global viral surveillance, more completely characterize novel viruses in reservoir host populations, and generate critical baseline knowledge to help prevent viral spillover.

REFERENCES

- [1] Olival KJ, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, et al. (2017) *Host and viral traits predict zoonotic spillover from mammals*. *Nature* 546: 646–650. doi: [10.1038/nature22975](https://doi.org/10.1038/nature22975)
- [2] Carroll D, Daszak P, Wolfe ND, Gao GF, Morel CM, et al. (2018) *The Global Virome Project*. *Science* 359: 872–874. doi: [10.1126/science.aap7463](https://doi.org/10.1126/science.aap7463)
- [3] Chao A, Chiu C-H. (2016) *Species Richness: Estimation and Comparison*. Wiley StatsRef: Statistics Reference Online. 1–26. doi: [10.1002/9781118445112.stat03432.pub2](https://doi.org/10.1002/9781118445112.stat03432.pub2)
- [4] Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrel CM, et al. (2013) *A strategy to estimate unknown viral diversity in mammals*. *mBio* 4: e00598-13. doi: [10.1128/mBio.00598-13](https://doi.org/10.1128/mBio.00598-13)
- [5] Anthony SJ, Islam A, Johnson C, Navarrete-Macias I, Liang E, et al. (2015) *Non-random patterns in viral diversity*. *Nature Communications* 6: 8147. doi: [10.1038/ncomms9147](https://doi.org/10.1038/ncomms9147)