**Accounting for imperfect detection reveals role of host traits in structuring viral diversity
of a wild bat community**

Anna R. Sjodin^{1,2,*}, Simon J. Anthony³, Michael R. Willig^{1,4}, Morgan W. Tingley^{1,5}

¹ Ecology and Evolutionary Biology, University of Connecticut, 75 N. Eagleville Road, Unit
3043, Storrs, Connecticut, USA 06269

² Biological Sciences, University of Idaho, 875 Perimeter Drive, MS 3051, Moscow, Idaho, USA
83844

³ Center for Infection and Immunity, Columbia University, 722 W. 168th Street, 17th Floor, New
York, New York, USA 10032

⁴ Center for Environmental Sciences & Engineering and Institute of the Environment, University
of Connecticut, Storrs, Connecticut, USA 06269

⁵ Ecology and Evolutionary Biology, University of California – Los Angeles, 621 Charles E
Young Drive South, Los Angeles, California, USA 90095

* Corresponding author

email: annasjodin03@gmail.com

Abstract: Understanding how multi-scale host heterogeneity affects viral community assembly can illuminate ecological drivers of infection and host-switching. Yet, such studies are hindered by imperfect viral detection. To address this issue, we used a community occupancy model – refashioned for the hierarchical nature of molecular-detection methods – to account for failed detection when examining how individual-level host traits affect herpesvirus richness in eight species of wild bats. We then used model predictions to examine the role of host sex and species identity on viral diversity at the levels of host individual, population, and community. Results demonstrate that cPCR and viral sequencing failed to perfectly detect viral presence. Nevertheless, model estimates correcting for imperfect detection show that reproductively active bats, especially reproductively active females, have significantly higher viral richness, and host sex and species identity interact to affect viral richness. Further, host sex significantly affects viral turnover across host populations, as females host more heterogeneous viral communities than do males. Results suggest models of viral ecology benefit from integration of multi-scale host factors, with implications for bat conservation and epidemiology. Furthermore, by accounting for imperfect detection in laboratory assays, we demonstrate how statistical models developed for other purposes hold promising possibilities for molecular and epidemiological applications.

Key Words:

cPCR, Chiroptera, Herpesvirus, Occupancy Modeling, Virus Community

1. Introduction:

Emerging infectious viruses, such as SARS-CoV 2 or Zika, represent single pathogen species embedded within multi-host, multi-pathogen systems. Community ecology – at the scale of host communities, and also at the scale of viral communities – is increasingly recognized as a productive framework with which to dissect such complex dynamics, illuminating risk factors associated with pathogen transmission, host switching, and spillover from wildlife to humans, among others [1, 2, 3, 4, 5, 6]. However, viral community ecology is limited by two major quantitative obstacles. First, most viruses are rare [7, 8, 9, 10, 11]. This means that large numbers of host individuals must be sampled for population-level detection, and resulting datasets are highly zero-inflated. Second, many wildlife viruses remain unknown to science [12, 13].

Currently, two of the most common methods for characterizing viral communities and finding new viral taxa are unbiased high-throughput sequencing (HTS; e.g. metagenomics) and consensus polymerase chain reaction (cPCR; [10]). HTS is advantageous when the goal is to construct an entire viral genome or when so little is known about the target taxon that primers cannot be designed. However, HTS lacks the sensitivity of PCR methods. cPCR is a method that relies on the use of degenerate primers for the broad amplification of both known and unknown viruses [10, 12]. It targets genome regions that are highly conserved among related viruses (e.g. at the family or genus level) and can detect multiple coinfecting viruses, if present. The degenerate primers bind with different efficiency to different viruses within a targeted taxon, so viruses with lower homology in the primer-binding region will not be detected as efficiently as would viruses with high sequence homology. Additionally, degenerate primers are, by definition, less specific. This reduces the sensitivity of cPCR and increases the probability of false negatives when viral load in the sample is low. This is a particular problem in wildlife surveys, where

many samples are taken from healthy or asymptomatic animals [4, 14, 15, 16, 17, 18], and absence of symptoms is often associated with low viral load [19, 20, 21]. As a result, both HTS and cPCR methods have a high possibility of failed detection of true infections (i.e. false negatives).

Viral ecological studies using HTS and cPCR methods have mostly failed to account for the issues of false negatives. In general, high rates of false negatives are acknowledged (e.g. [4, 22]) but are rarely considered a major problem preventing honest inference. This becomes particularly problematic for studies of viral diversity and community assembly when accurate measures of the number and identity of detected viruses are necessary to infer patterns and mechanisms of co-infecting viruses. As consequent viral OTUs are missed for each sampled individual, the missing data compound when individuals are aggregated at the population and community levels. A handful of recent studies have used occupancy modeling techniques, adapted from wildlife ecology (where sites are surveyed to detect species, rather than where individual animals are assayed to detect viral strains), to better account for imperfect detection in an epidemiological context [23, 24, 25, 26]. However, as best we are aware, no one has yet adapted these methods to the detection process when using HTS or cPCR for biodiversity discovery or characterization of viral community ecology [27], and it is not a common practice to statistically account for failed detection in such a way across molecular biology more generally.

Additionally, viral ecologists have struggled to synthesize community dynamics across host scales. For example, infection heterogeneity among individuals is a key element in infectious disease dynamics [28, 29, 30] that is still relatively unexplored in viral communities of wildlife. Pathogen aggregation in host individuals can affect pathogen interactions, transmission, and host fitness, scaling up to alter population and community dynamics of the host. Failure to

account for such heterogeneity has potentially dangerous or expensive consequences for modelling outbreak probabilities and forecasting disease spread [31, 32, 33].

Bats are an important host to understand from the perspective of viral transmission, as they seem to harbor a high number of viruses capable of infecting humans [34, 35, 36]. Cave-roosting bats in Puerto Rico are abundant and relatively easily captured, making them a practical system for addressing viral dynamics from an ecological perspective. Additionally, much is known about their behavior, biology, and life history that can guide contextual understanding of viral community dynamics [37]. For example, most cave-roosting species in Puerto Rico show sex-related spatial segregation during reproduction [37]. Females often use a centralized maternity colony within the cave, where they frequently and repeatedly come into contact with the same individuals. Males do not use this space. By continually re-exposing themselves to the same individual conspecifics, females likely reduce the variety of viruses to which they are exposed. Thus, host sex likely effects pathogen diversity. Indeed, the role of host sex has shown female-biased effects on ectoparasite and macro-parasite infection in Puerto Rican bats [38, 39]. Furthermore, many Puerto Rican caves are well-studied in terms of bat diversity and abundance. Two adjacent caves – Culebrones and Larva – have been studied extensively [37, 38, 39, 40, 41, 42, 43, 44, 45]. Species composition, as well as population sizes, diets, and reproductive phenology are all well understood for the bats roosting in these caves.

The goal of this study was to examine the role of host heterogeneity on viral diversity, while accounting for rarity and failed detection. To address these objectives, we collected wild bats from Culebrones and Larva Caves in Puerto Rico, tested bats for herpesvirus infection, and recorded individual-level host traits (mass, forearm length, ectoparasite intensity, sex, and reproductive status). To examine viral diversity across the scales of individual and population,

we quantified α (individual-level) and γ (population-level) viral diversity, from which we calculated β diversity (compositional turnover of viruses among individual hosts).

Our research aimed to test multiple hypotheses in our inferential framework. First, we hypothesized that larger host individuals would have higher viral richness, as larger hosts provide more habitat space for pathogens and have shown higher levels of infection in previous work [46]. Second, we hypothesized that bats with more ectoparasites would harbor more viruses, as high parasite load and failure to groom have been shown previously to indicate illness or lowered immunity of hosts [47, 48, 49]. Third, we hypothesized that females and reproductively active individuals would have higher viral richness, as host sex [38, 39] and reproductive status [46, 50] have demonstrated impacts on viral infection in a diversity of host systems. Fourth, we hypothesized that bats roosting in the larger cave would have higher viral richness, due to greater host diversity and abundance in the larger cave. Relating to the community composition and diversity of infections, fifth, we hypothesized that β diversity of viruses would be greater in male than in female bats, as females limit their habitat range and use maternity colonies extensively during reproductive season, homogenizing their viral communities. Driven by relationships at the α and β levels, sixth, we hypothesized that γ diversity would be higher in female bats than male bats. Seventh, we hypothesized that γ diversity would be significantly different among host species, with species representing larger host populations having higher γ diversity than those roosting in smaller numbers. Finally, driven by strong relationships at the α and γ levels, we expected β diversity to differ among host species.

2. Materials and Methods:

(a) *Sample Collection*

Field work was conducted at Mata de Plátano Nature Reserve in Arecibo, Puerto Rico (18° 24.868' N, 66° 43.531' W). Mata de Plátano Nature Reserve (operated by InterAmerican University, Bayamon, Puerto Rico) is in the north-central, karstic region of Puerto Rico, an area dominated by caves suitable for roosting bats. Samples were collected from apparently healthy bats in Culebrones and Larva Caves for a total of 35 nights between June and August of 2017. Details about the cave habitats can be found in the Supplemental Material.

A harp trap, mist nets, and hand nets were used to capture 1,086 bats representing eight species: *Pteronotus quadridens*, *P. portoricensis*, *Mormoops blainvillii*, *Eptesicus fuscus*, *Monophyllus redmani*, *Brachyphylla cavernarum*, *Artibeus jamaicensis*, and *Erophylla sezekorni*. Upon capture, each bat was placed into a cotton holding bag, and individuals were identified to species based on Gannon et al. [37]. A clean cotton-tipped swab was used to collect saliva from each bat's mouth. Swabs were placed in individual cryovials containing viral transport medium, maintained at -80° C in a dry shipper, and sent to Columbia University's Center for Infection and Immunity. From each captured bat, we recorded six traits: mass, forearm length, ectoparasite intensity, sex, and reproductive status. Detailed descriptions of trait measurements can be found in the Appendix. Additionally, wing punches were taken from each bat for a separate analyses. To prevent re-sampling of recaptured bats, all subsequently captured bats with a punch-sized hole in the wing were released. All methods were approved by the University of Connecticut Institutional Animal Care and Use Committee (IACUC, protocol A15-032).

(b) Laboratory Analyses

For each sample, herpesvirus cPCR assay targeting a ~170 base pair region of the catalytic subunit of the DNA polymerase gene of the herpesvirus family [51] was run twice (Figure 1). For assay one, all cPCR gel products (1% agarose) of expected size (Figure 1, b1)

were cloned into Strataclone PCR cloning vector (Figure 1, c1), and twelve colonies were sequenced to detect viral co-occurrence and confirm viral identity via cross-referencing against the GenBank nucleotide database (Figure 1, d1). No cut bands from assay one resulted in false positives, so for assay two, only those gel products of expected size that were *not* present in assay one (Figure 1, b2) were cloned (Figure 1, c2) and sequenced (Figure 1, d2). Consequently, there were multiple samples from assay two for which data exist only on herpesvirus detection (positive or negative), but not on the identity of the detected herpesviruses (Figure 1, e2). Operational taxonomic units (OTUs) were determined using percent nucleotide identity and affinity propagation [11, 52] and were used in place of viral species.

Each sample was tested for PCR inhibition using TaqMan Exogenous Internal Positive Control Reagents (Applied Biosystems) with quantitative PCR as described in the manufacturer's instruction manual. Amount of inhibition was determined using the standardized difference of sample quantitation cycle (Cq) from the mean Cq of two negative controls.

(c) *Modeling Framework*

To explore the effects of individual-level host traits on herpesvirus infection, we fit a Bayesian community-level occupancy model to OTU-host occurrence data [53, 54]. This framework explicitly differentiates between the underlying state process determining the presence or absence of each OTU and the observation process determining the detection or non-detection of each OTU, conditional on its true presence. Although occupancy models have been employed in the field of animal ecology to study the occurrence of vertebrates across a landscape, the modeling framework has rarely yet been applied in molecular settings to account for imperfect detection during indirect sampling, such as with eDNA [55, 56] or in studies of pathogen ecology [23, 24, 25].

By modeling OTUs in a community context, we improved accuracy of estimates of incidence for rare OTUs by considering parameter estimates as all coming from a similar stochastic process, a Bayesian process known as “borrowing strength” [57, 58]. As a trade-off, the model does not robustly inform which host traits shape occupancy of individual OTUs, but provides strong inference on host trait relationships to the aggregate community of OTUs.

In the model, response variable y_{ijk} is a binomial random variable denoting whether herpesvirus OTU i was detected ($y_{ijk} = 1$) or not ($y_{ijk} = 0$) in host individual j during assay k (green values in Figure 1e and f). A second response variable, Y_{jk} , is a binomial random variable denoting whether any herpesvirus, regardless of OTU identity, was detected ($Y_{jk} = 1$) or not ($Y_{jk} = 0$) in host individual j during assay k (purple values in Figure 1 e and f). The two response variables, y and Y , are hierarchically related, yet independently observed via previously described methods: cPCR identifies the presence or non-detection of any herpesvirus (Y), and subsequent sequencing and bioinformatics identifies the presence or non-detection of particular OTUs (y).

The model simultaneously estimates the probability of infection (ψ_i) and the probability of detection (p_{ijk}) with a mixture model in the form $y_{ijk} \sim \text{Bernoulli}(p_{ijk} * z_{ij})$, where p_{ijk} is the probability of detecting OTU i in host individual j during assay k , and z_{ij} is a latent variable representing true infection of OTU i in host individual j , deriving from the equation $z_{ij} \sim \text{Bernoulli}(\psi_i)$. A second latent variable (Z_j) represents true herpesvirus infection, regardless of OTU identity, in host individual j . The model’s hierarchical structure allows the two response variables (y_{ijk} and Y_{jk}), as well as the two latent variables (z_{ij} , and Z_j), to inform one another, as $Z_j = \max(z_{ij})$ and $Y_{jk} \sim \text{Bernoulli}(p_{\text{assay}} * Z_j)$, where p_{assay} represents a constant probability that the cPCR assay will detect a herpesvirus given the presence of at least one OTU. Together, the model

assumes that an empirical detection ($y_{ijk} = 1$, $Y_{jk} = 1$) represents true viral infection, but that an empirical non-detection ($y_{ijk} = 0$, $Y_{jk} = 0$) could indicate either a true absence or a true infection with failed detection.

To structure the model, sample inhibition and OTU subfamily were included as fixed effects influencing OTU-specific detection probability (p_{ijk}). Host forearm length, cave identity, ectoparasite intensity, sex, reproductive status, and the interaction between sex and reproductive status were included as fixed effects for infection probability (ψ_{ij}). Host species was included as a random intercept for both detection and infection probabilities.

As is typical for community occupancy models, every intercept and slope parameter for a given herpesvirus was assumed to be drawn from a hierarchical Normal distribution that describes the distribution of slopes for all herpesvirus species available for sampling [59]. In this way, primary inference on the community is made by the examination of the hyper-parameters that provide a mean community-wide effect for any given covariate. As preliminary explorations of the dataset indicated that some herpesviruses were unusually common or unusually rare, instead of using a hierarchical Normal to describe the random distribution of occupancy intercepts, we used a fat-tailed distribution, the t -distribution. Full model code is provided online as Supplemental Material.

We executed the model in JAGS [60] using the R2jags package [61] with vague priors (e.g., normal with $\mu = 0$, $\tau = 0.1$, gamma with shape and rate = 0.01). We ran three chains of 15,000 iterations thinned by 50 with a burn-in of 25,000, yielding a posterior sample of 900 iterations across all chains. We checked convergence visually with traceplots and made parameter inference based on 95% Bayesian credible intervals. Posterior predictive checks were conducted for individual OTUs by calculating Bayesian p-values [62] and 90% Bayesian

credible intervals for mean detected prevalence of each OTU. A lack of fit for a particular OTU would be indicated by a 90% credible interval that did not overlap the empirically observed prevalence. All analyses were done using R version 3.4.3 [63].

(d) Alpha, Beta, and Gamma Partitions

To examine viral diversity across the levels of host individual and population, we compared α , β , and γ diversity among host species and between host sexes using 900 host-by-OTU matrices drawn from the community occupancy model's posterior distribution of the true underlying infection state, $z_{i,j}$. α diversity was calculated as the average viral richness per individual bat across 1) all species for each sex, 2) both sexes for each species, and 3) each species-by-sex combination (male *A. jamaicensis*, female *A. jamaicensis*, male *P. quadridens*, etc.). γ diversity was defined as the total viral richness for 1-3. β diversity was calculated using both the multiplicative ($\gamma = \alpha * \beta$; [64]) and additive ($\gamma = \alpha + \beta$; [65]) models.

For α diversity, the general statistical design is that of a two-factor analysis of variance (ANOVA) with replication, with individual bats representing replicates. To test the significance of results, we calculated F-statistics for differences in α diversity associated with 1-3. We then compared empirical F-statistics for each of the 900 posterior draws to three null distributions (H_0 : no difference in α diversity between host sexes or among host species). Each null distribution was generated using a different assumptions regarding the relationship between relative abundances of each sex-by-species group, as compared to the true proportion of individuals in that group. Details describing generation of the null distribution are available in the Appendix.

For β and γ diversity, the general statistical design is that of a two-factor ANOVA without replication, with treatment groups representing replicates. As such, values were only compared between sexes and among species (1 and 2). We could not assess their interaction.

Simulation methods were identical to those for the α -level analyses (Appendix). All analyses were done using R version 3.4.3 [63]. R code used for the simulation methods and calculation of F-statistics is available at github.com/ARSjodin/AlphaBetaGamma.

3. Results:

(a) *Model Fit*

Screening results and OTU delineations are reported previously [11]. Posterior predictive checks on each individual OTU indicate an overall adequate degree of model fit, with no 90% credible intervals falling outside of the observed prevalence for an OTU (Supplemental Table 2). Observed prevalence was more likely to be underpredicted than overpredicted, indicating that the model may be more pessimistic about detection probability than observed in the real dataset.

(b) *Imperfect detection of OTUs*

Model fits indicated OTUs were imperfectly detected at all stages of laboratory analysis. The cPCR assay had a 0.778 probability (p_{assay} , 95CrI = 0.744–0.807) of detecting a herpesvirus given the presence of at least one OTU (Table 1). Given two cPCR assays, this suggests a 0.049 probability (95CrI = 0.037–0.065) of incorrectly assuming a host is not infected, were detection assumed to be perfect. However, even if the cPCR assay suggests a positive infection, not every OTU will be detected via sequence amplification. On average across all OTUs, the probability of detecting a single OTU via sequencing was 0.699 (inverse logit-transformed μ_{a0} , 95CrI = 0.637–0.762). There was no support for a role of either sample inhibition or viral subfamily in affecting viral detection rates (Table 1). Combined across both stages of viral detection, there was both a non-zero probability that cPCR assays would fail to detect a viral infection, and that sequence amplification would then fail to identify the full suite of OTUs present. Consequently, the fitted model estimated on average that 1.75 hosts (95CrI = 0 – 5) likely held infections of at least one

OTU that were not detected by cPCR, and that the mean number of OTUs missed per host was 0.036 (95CrI = 0.017–0.063). Across the tested host community (n = 41), this combined to an expected undetected total of 12.2 OTU infections (95CrI = 5.5–21.0).

(c) *Role of Host Traits*

In general, herpesvirus occupancy was low in Puerto Rican bats, but the model uncovered robust support for the importance of individual-level traits as drivers of viral richness (Table 1). Herpesvirus richness was higher in reproductively active bats compared to non-reproductive bats. Specifically, reproductively active female bats had higher viral richness as compared to both non-reproductive bats and to reproductively-active male bats. No other host traits significantly affected the probability of viral occurrence.

Using model estimates of true occurrence correcting for imperfect detection, bats showed differing degrees of herpesvirus diversity across community scales. α diversity was significantly different between the sexes, but this effect differed by species, with only some bat species showing sex-specific differences in α diversity (Figure 2, Appendix Table 3). Despite strong differences in α diversity by species and sex, this did not scale up to γ -level differences (Figure 2, Appendix Table 3). Taken together, this means that although certain bat species (and sexes within those species) host more OTUs than do other species, the identity of the OTUs is quite similar across individuals of a species, indicating low OTU turnover at the population level. Indeed, β diversity did not differ significantly across species (Figure 2, Appendix Table 3). However, the multiplicative form of β diversity (i.e. $\gamma = \alpha * \beta$) was significantly different between sexes, with higher β diversity, and therefore higher turnover of OTUs, in females than males (Figure 2, Appendix Table 3). Because no γ -level difference existed between bat sexes,

and the sex differences at the α level depended on species, the differences in β diversity between sexes is likely driven by α -level effects of certain species.

4. Discussion:

Scientists and public health officials have called for a shift from post-emergence, reactive viral studies to predictive, preventative approaches [66, 67]. However, the integration of disease ecology and ecological forecasting is still in its infancy and is constrained in utility [67, 68, 69, 70, 71, 72, 73]. An ongoing hindrance to progress is data latency, or the lag time between data collection and data availability for analyses [72]. In the context of understanding viral shedding in wildlife, this includes the time spent capturing and sampling hosts in the field, transporting samples from the field to the laboratory, extracting genetic material and diagnosing infection, and subsequent sequencing and bioinformatics. By linking increased viral shedding to more easily and quickly attainable host attributes such as reproduction and the interaction between sex and reproductive state, our study suggests that the use of host traits may serve as a proxy for transmission risk, providing a potential solution to the data latency problem. For example, in many bat species, including those tested here [37], reproductive status is a trait that is predictable based on season, making it a more useful proxy in forecasting [74]. Assuming this relationship holds for other virus-host systems, when scientists make forecasts about the risk posed by bats to human health, reproductive status could potentially be added to models and updated temporally without extensive, ongoing molecular sampling. Ongoing investigations the relationship between viruses and host reproduction would help illuminate the utility of this proxy.

Although all living organisms contain a multitude of viruses [34, 75, 76], targeted testing of bats, and the resulting diversity of discovered viruses, have created conflicting perceptions about the threat or value of bats to humans. Dramatized communication of bat surveillance has

exacerbated the perceived negative reputation of bats, causing viral discovery efforts to become a conservation issue [77]. Unfortunately, efforts to intentionally kill wild bats for protection of human health have threatened populations worldwide [78, 79, 80]. However, the effect of reproduction on viral shedding, as presented here, suggests non-contradictory goals of bat conservation and public health. Because bats, especially female bats, shed more viruses during reproduction, the risk of viral transmission from bats to humans is higher during this time. Additionally, reproductive rate is an intrinsic factor that informs bat conservation [81] and correlates with extinction rate [82]. Disturbances caused by humans during reproduction can increase mortality rates of offspring [84], which can then cause population declines and threaten population persistence. Therefore, minimizing human disturbance to bats during reproduction, especially to maternity colonies, can curtail the risk of viral transmissions to humans and human-induced declines of bat populations.

At the level of individual hosts (α), host sex and species interacted to significantly affect viral diversity (Table 2). This suggests that the difference in α diversity among the host species is different for males and females of each species. This could be a result of an antagonistic relationship between host species and sex (e.g. males of species one have *higher* viral diversity than females of species one, but males of species two have *lower* viral diversity than females of species two). Alternatively, this could be a result of differing magnitude of viral diversity among host species and sex (e.g. males of species one have *much* higher viral diversity than females of species one, but males of species two have only *slightly* higher viral diversity than females of species two). Values of α diversity (Figure 2, Appendix Table 3) suggest that the differences in viral richness between sexes generally differ in magnitude among host species, and the impact of sex is not antagonistic among species.

Host sex, alone, had a significant effect on β diversity (using the multiplicative model), with female bats hosting more compositionally heterogeneous infracommunities (i.e. higher β diversity; Table 2, Figure 2). This was contrary to our hypothesis and shows that, despite limiting range size and repeated exposure to the same individual conspecifics, the viral metacommunity of female bats does not become homogenized compositionally. Instead, compositional heterogeneity could be a result of strong physiological differences between immune-competence of the sexes, especially during reproduction. Immune suppression of reproductively active female bats may activate previously latent herpesviruses, resulting in viral infracommunities that comprise OTUs acquired at any point in time. However, this could not be tested in our study. If supported, it would suggest that viral composition in female bats may not be as temporally constrained as in males, resulting in a more compositionally heterogeneous metacommunity.

Notably, differences in β diversity related to host sex are shaped strongly by two host species, *P. quadridens* and *P. portoricensis* (Appendix Table 3). Unfortunately, little is known about behavioral or physiological differences between the sexes of each of these species, that may contribute to viral sharing in a manner that is different from that of other species. Investigations of the mechanistic relationship between host physiology and viral infection is a fruitful area for further research. Additionally, this difference in viral β diversity between male and female *P. quadridens* is due primarily to differences in α diversity between the sexes, as γ diversity of male and female *P. quadridens* are indistinguishable (Appendix Table 3). For *P. portoricensis*, the relationship is reversed. These results further emphasize the need to explicitly incorporate both individual- and population-level heterogeneity in analyses of viral ecology.

In terms of viral detection, neither sample inhibition nor OTU subfamily influenced overall viral detection rates (Table 1). The near-zero effect size of sample inhibition is important

because it demonstrates that samples do not contain contamination or non-viral genetic material that prohibited detection of herpesviruses. Additionally, the near-zero effect size of OTU subfamily shows that the assay is not biased towards either the Beta- or Gammaherpesvirinae. Notably, no alphaherpesviruses were detected in the samples [11], so the results make no suggestions about the assay's efficiency in detecting viruses belonging to this subfamily. Still, detection was imperfect, and reasons driving failed detection are yet to be determined. One unmeasured factor that may influence the probability of detection in this system is viral load, or the intensity of viral infection (i.e. viral "abundance" in each host). Indeed, infection intensity is a significant driver of detection probability in other disease systems [24, 25]. In the context of the present study, however, quantifying viral intensity in each sample would require development of separate quantitative PCR assays for each OTU and was beyond the scope of this work.

Failed detection in this system can occur at the stage of sample collection (i.e. the host is infected, but the virus is not in the sample) and at the stage of molecular analyses (i.e. the virus is in the sample, but the laboratory methods did not detect it). Reasons could be biological (e.g. low viral load associated with asymptomatic hosts) or methodological, if tests fail because of genetic variation in the primer-binding sites or some other mechanism, such as failure to capture low-level sequences in the cloning step. The methods used here only account for failed detection at the stage of molecular analyses – as only one sample was tested per host – thus failed detection at the stage of sample collection will be perceived here as a lack of occurrence. However, our modeling method could easily be expanded to accommodate the former, additional source of heterogeneity, by taking and testing multiple samples per host.

Our study illustrates how viral community ecology can be used to identify risk factors for infection while quantitatively accounting for imperfect detection, serving as a model for studies of viral communities moving forward. Community-level occupancy modeling can be applied to continued viral discovery efforts, molecular biology more broadly, and to applications of approaches or concepts from community ecology to virus systems. For the herpesvirus assay used in this study [51], the model indicated that two assay runs were able to detect approximately 95% of true infections. A third run would have resulted in an estimated 99% detection. Similar calculations, following [84], can be done for other assays to guide the design of molecular methods. Additionally, the development of a novel hierarchical model that uses a single assay run with specific OTU identification (Figure 1, e1), followed by assay runs that detect infection only, without OTU identification (Figure 1, e2), accounts for false absences without considerably increasing sequencing costs or time spent analyzing additional sequences. The further co-development of emerging quantitative models with molecular techniques stands to improve both the strength of findings as well as the cost-efficiency of assays.

Acknowledgements: We thank Armando Rodríguez-Duran for logistical help in the field, as well as the field team, including Sarah Stankavich, Jose Rivera, Melanie Hodge, Brian Springall, Elspeth Pierce, Emily Stanford, Juan Contreras, Ismael Ribot, and Frank Sjodin. Isamara Navarrete-Macias and Eliza Liang trained ARS on laboratory methods, and Heather Wells provided lab and bioinformatics help. Eliza Grames provided invaluable coding assistance. We also thank Janine Caira, Sarah Knutie, and Mark Urban for constructive discussions that advanced the intellectual development of the work. A.R.S. was funded by the NSF Graduate Research Fellowship Program and a Jorgensen Fellowship from the University of Connecticut. Both A.R.S. and M.W.T. additionally received support from The Center for Biological Risk at

the University of Connecticut. M.R.W. was funded by a grant from NSF (DEB-1546686 and DEB-1831952). Fieldwork was supported by Sigma Xi; the Royal Society of Tropical Medicine and Hygiene; and the University of Connecticut's Department of Ecology and Evolutionary Biology, El Instituto via the Tinker Foundation, and the Office of the Vice President for Research. This project also benefited from support from the USAID PREDICT project.

Author Contributions: A.R.S. conducted field and laboratory work. S.J.A. and A.R.S. designed laboratory methods. M.W.T. and A.R.S. designed modeling methods. M.R.W. and A.R.S. designed simulation methods. M.W.T. and A.R.S. conducted quantitative analyses and drafted the manuscript. All authors provided comments and gave final approval for publication.

References:

1. Johnson PTJ, deRoode JC, Fenton A. 2015 Why infectious disease research needs community ecology. *Science* **349**, 1259504. (doi: 10.1126/science.1259504)
2. Seabloom EW, Borer ET, Gross K, Kendig AE, Lacroix C, Mitchell CE, Mordecai EA, Power AG. 2015 The community ecology of pathogens: coinfection, coexistence and community composition. *Ecol. Lett.* **18**, 401-415. (doi: 10.1111/ele.12418)
3. Díaz-Muñoz SL. 2017 Viral coinfection is shaped by host ecology and virus-virus interactions across diverse microbial taxa and environments. *Vir. Evol.* **3**, 1-15. (doi: 10.1093/ve/vex011)

4. Escalera-Zamudio M, Taboada B, Rojas-Anaya E, Löber U, Loza-Rubio E, Arias CF, Greenwood AD. 2017 Viral communities among sympatric vampire bats and cattle. *EcoHealth*. **15**, 132-142. (doi: 10.1007/s10393-017-1297-7)
5. Fountain-Jones NM, Packer C, Troyer JL, VanderWaal K, Robinson S, Jacquot M, Craft ME. 2017 Linking social and spatial networks to viral community phylogenetics reveals subtype-specific transmission dynamics in African lions. *J. Anim. Ecol.* **86**, 1469-1482. (doi: 10.1111/1365-2656.12751)
6. Bergner LM, Orton RJ, Benavides JA, Becker DJ, Tello C, Biek R, Streicker DG. 2019 Demographic and environmental drivers of metagenomics viral diversity in vampire bats. *Mol. Ecol.* **29**, 26-39. (doi: 10.1111/mec.15250)
7. Tang XC, Zhang JX, Zhang SY, Wang P, Fan XH, Li LF, Li G, Dong BQ, Liu W, Cheung CL, et al. 2006 Prevalence and genetic diversity of coronaviruses in bats from China. *J. Virol.* **80**, 7481-7490. (doi: 10.1128/JVI.00697-06)
8. Winker K, Spackman E, Swayne DE. 2008 Rarity of Influenza A virus in spring shorebirds, southern Alaska. *Emerg. Infect. Dis.* **14**, 1314-1316. (doi: 10.3201/eid1408.080083)
9. Wacharapluesadee S, Boongrid K, Wanghonsa S, Ratanasetyuth N, Supavongwong P, Saengsen D, Gongal GN, Hemachudha T. 2010 A longitudinal study of the prevalence of Nipah virus in *Pteropus lylei* bats in Thailand: evidence for seasonal preference in

- disease transmission. *Vector Borne Zoonotic Dis.* **10**, 183-190. (doi: 10.1089/vbz.2008.0105)
10. Anthony SJ, Islam A, Johnson C, Navarrete-Macias I, Liang E, Jain K, Hitchens PL, Che X, Solovyov A, Hicks AL, et al. 2015 Non-random patterns in viral diversity. *Nature Comm.* **6**, 8147. (doi: 10.1038/ncomms9147)
11. Sjodin AR, Willig MR, Anthony SJ. 2019 Quantitative delineation of herpesviruses in bats for use in ecological studies. *BioRxiv.* (doi: 10.1101/856518)
12. Anthony SJ, Epstein JH, Murray KA, Navarrete-Macias I, Zambrana-Torrel CM, Solovyov A, Ojeda-Flores R, Arrigo NC, Islam A, Khan SA, et al. 2013 A strategy to estimate unknown viral diversity in mammals. *mBio* **4**, e00598-13. (doi: 10.1128/mBio.00598-13)
13. Carlson, C, Zipfel CM, Garnier R, Bansal S. 2019 Global estimates of mammalian viral diversity accounting for host sharing. *Nature Ecol. Evol.* **3**, 1070-1075. (doi: 10.1038/s41559-019-0910-6)
14. Aguilar-Setien A, Loza-Rubio E, Salas-Rojas M, Brisseau N, Cliquet F, Pastoret PP, Rojas-Dotor S, Tesoro E, Kretschmer R. 2005 Salivary excretion of rabies virus by healthy vampire bats. *Epidemiol. Infect.* **133**, 517-522. (doi: 10.1017/S0950268805003705)

15. Savic V, Labrovic A, Zelenika TA, Balenovic M, Separovic S, Jurinovic L. 2010 Multiple introduction of Asian H5N1 avian influenza virus in Croatia by wild birds during 2005-2006 and isolation of the virus from apparently healthy black-headed gulls (*Larus ridibundus*). *Vector Borne Zoonotic Dis.* **10**, 915-920. (doi: 10.1089/vbz.2009.0107)
16. Tse H, Tsang AKL, Tsoi HW, Leung ASP, Ho CC, Lau SKP, Woo PCY, Yuen KY. 2012. Identification of a novel bat papillomavirus by metagenomics. *PLoS One.* **7**, e43986. (doi: 10.1371/journal.pone.0043986)
17. Wray AK, Olival KJ, Morán D, Lopez MR, Alvarez D, Navarrete-Macias I, Liang E, Simmons NB, Lipkin WI, Daszak P, Anthony SJ. 2016 Viral diversity, prey preference, and *Bartonella* prevalence in *Desmodus rotundus* in Guatemala. *EcoHealth.* **13**, 761-774. (doi: 10.1007/s10393-016-1183-z)
18. Plowright RK, Peel AJ, Streicker DG, Gilbert AT, McCallum H, Wood J, Baker ML, Restif O. 2017 Transmission or within-host dynamics driving pulses of zoonotic viruses in reservoir-host populations. *PLoS Negl. Trop. Dis.* **10**, e0004796. (doi: 10.1371/journal.pntd.0004796)
19. Granados A, Goodall EC, Luinstra K, Smieja M, Majony J. 2015 Comparison of asymptomatic and symptomatic rhinovirus infections in university students: incidence, species diversity, and viral load. *Diagn. Microbiol. Infect. Dis.* **82**, 292-296. (doi: 10.1016/j.diagmicrobio.2015.05.001)

- 497
- 498 20. Zhou W, Ullman K, Chowdry V, Reining M, Benyeda Z, Baule C, Juremalm M, Wallgren P,
499 Schwarz L, Zhou E, Pina Pedrero S, Hennig-Pauka I, Segales J, Liu L. 2016 Molecular
500 investigations on the prevalence and viral load of enteric viruses in pigs from five
501 European countries. *Vet. Microbiol.* **182**, 75-81. (doi: 10.1016/j.vetmic.2015.10.019)
502
- 503 21. Brettell LE, Mordecai GJ, Schroeder DC, Jones IM, da Silva JR, Vicente-Rubiano M, Martin
504 SJ. 2017 A comparison of deformed wing virus in deformed and asymptomatic honey
505 bees. *Insects.* **8**, 1-12. (doi: 10.3390/insects8010028)
506
- 507 22. McClintock BT, Nichols JD, Bailey LL, MacKenzie DI, Kendall WL, Franklin AB. 2010
508 Seeking a second opinion: uncertainty in disease ecology. *Ecol. Lett.* **13**, 659-674. (doi:
509 10.1111/j.1461-0248.2010.01472.x)
510
- 511 23. Lachish S, Gopalaswamy AM, Knowles SCL, Sheldon BC. 2011 Site-occupancy modelling
512 as a novel framework for assessing test sensitivity and estimating wildlife disease
513 prevalence from imperfect diagnostic tests. *Methods Ecol. Evol.* **3**, 339-348. (doi:
514 10.1111/j.2041-210X.2011.00156.x)
515
- 516 24. Miller DAW, Talley BL, Lips KR, Grant EHC. 2012 Estimating patterns and drivers of
517 infection prevalence and intensity when detection is imperfect and sampling error occurs.
518 *Methods Ecol. Evol.* **3**, 850-859. (doi: 10.1111/j.2041-210X.2012.00216.x)
519

25. DiRenzo, GV, Campbell Grant EH, Longo AV, Che-Castaldo C, Zamudio KR, Lips KR. 2017 Imperfect pathogen detection from non-invasive skin swabs biases disease inference. *Methods Ecol. Evol.* **9**, 380-389. (doi: 10.1111/2041-210X.12868)
26. Tabak MA, Pedersen K, Miller RS. 2019 Detection error influences both temporal seroprevalence predictions and risk factors associations in wildlife disease models. *Ecol. Evol.* **9**, 10404-10414. (doi: 10.1002/ece3.5558)
27. Bogich TL, Anthony SJ, Nichols JD. 2013 Surveillance theory applied to virus detection: a case for targeted discovery. *Future Virol.* **8**, 1201-1206. (doi: 10.2217/fvl.13.105)
28. Paull SH, Song S, McClure KM, Sackett LC, Kilpatrick AM, Johnson PTJ. 2012 From superspreaders to disease hotspots: linking transmission across hosts and space. *Front. Ecol. Evol.* **10**, 75-82. (doi: 10.1890/110111)
29. VanderWaal KL, Ezenwa VO. 2016 Heterogeneity in pathogen transmission: mechanisms and methodology. *Funct. Ecol.* **30**, 1606-1622. (doi: 10.1111/1365-2435.12645)
30. Stephenson JF, Young KA, Fox J, Jokela J, Cable J, Perkins SE. 2017 Host heterogeneity affects both parasite transmission to and fitness on subsequent hosts. *Phil. Trans. R. Soc. B.* **372**, 20160093. (doi: 10.1098/rstb.2016.0093)

31. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. 2005 Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355-359. (doi: 10.1038/nature04153)
32. Yates A, Antia R, Regoes RR. 2006 How do pathogen evolution and host heterogeneity interact in disease emergence? *Proc. R. Soc. B* **273**, 3075-3083. (doi: 10.1098/rspb.2006.3681)
33. Gog JR, Pellis L, Wood JLN, McLean AR, Arinaminpathy N, Lloyd-Smith JO. 2015 Seven challenges in modeling pathogen dynamics within-host and across scales. *Epidemics* **10**, 45-48. (doi: 10.1016/j.epidem.2014.09.009)
34. Olival JK, Hosseini PR, Zambrana-Torrel C, Ross N, Bogich TL, Daszak P. 2017 Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646-650. (doi: 10.1038/nature22975)
35. Han BA, Kramer AM, Drake JM. 2016 Global patterns of zoonotic disease in mammals. *Trends Parasitol.* **32**, 565-577. (doi: 10.1016/j.pt.2016.04.007)
36. Brook CE, Dobson AP. 2015 Bats as 'special' reservoirs for emerging zoonotic pathogens. *Trends Microbiol.* **23**, 172-180. (doi: 10.1016/j.tim.2014.12.004)

37. Gannon MR, Kurta A, Rodríguez-Duran A, Willig MR. 2005 *Bats of Puerto Rico*. Lubbock, TX: Texas Tech University Press.
38. Kurta A, Whitaker Jr. JO, Wrenn WJ, Soto-Centeno JA. 2007 Ectoparasitic assemblages on mormoopid bats (Chiroptera: Mormoopidae) from Puerto Rico. *J. Med. Entomol.* **44**, 953-958. (doi: 10.1093/jmedent/44.6.953)
39. Krichbaum K, Perkins S, Gannon MR. 2009 Host-parasite interactions of tropical bats in Puerto Rico. *Acta Chiropt.* **11**, 157-162. (doi: 10.3161/150811009X465776)
40. Rodríguez-Duran A. 1998 Nonrandom aggregations and distribution of cave-dwelling bats in Puerto Rico. *J. Mammal.* **79**, 141-146. (doi: 10.2307/1382848)
41. Jones KE, Barlow KE, Vaughan N, Rodríguez-Duran A, Gannon MR. 2001 Short-term impacts of extreme environmental disturbance on the bats of Puerto Rico. *Anim. Conserv.* **4**, 59-66. (doi: 10.1017/S1367943001001068)
42. Soto-Centeno JA, Kurta A. 2006 Diet of two nectarivorous bats, *Erophylla sezekorni* and *Monophyllus redmani* (Phyllostomidae) on Puerto Rico. *J. Mammal.* **87**, 19-26. (doi: 10.1644/05-MAMM-041R1.1)
43. Dittmar K, Mayberry JR. 2010 Bat activity in large roosts drives diurnal cave microclimate variation. *Speleobiology Notes* **2**, 12-14. (doi: 10.5563/spbn.v2i0.23)

44. Olival KJ, Dittmar K, Bai Y, Rostal MK, Lei BR, Daszak P, Kosoy M. 2015 *Bartonella* spp. in a Puerto Rican bat community. *J. Wildl. Dis.* **51**, 274-278. (doi: 10.7589/2014-04-113)
45. Rolfe AK, Kurta A, Clemans DL. 2014 Species-level analysis of diets of two mormoopid bats from Puerto Rico. *J. Mammal.* **95**, 587-596. (doi: 10.1644/13-MAMM-A-190)
46. Plowright RK, Field HE, Smith C, Divljan A, Palmer C, Tabor G, Daszak P, Foley JE. 2008 Reproduction and nutritional stress are risk factors for Hendra virus infection in little red flying foxes (*Pteropus scapulatus*). *Proc. R. Soc. B* **275**, 861-869. (doi: 10.1098/rspb.2007.1260)
47. Hart BL. 1988 Biological basis of the behavior of sick animals. *Neurosci. Biobehav. Rev.* **12**, 123-137. (doi: 10.1016/S0149-7634(88)80004-6)
48. Hart BL. 2010 Beyond fever: comparative perspectives on sickness behavior. In *Encyclopedia of Animal Behavior*, Vol. 1 (eds M Breed and J Moore) Oxford, UK: Academic Press.
49. Stockmaier S, Bolnick DI, Page RA, Carter GG. 2018 An immune challenge reduces social grooming in vampire bats. *Anim. Behav.* **140**, 141-149. (doi: 10.1016/j.anbehav.2018.04.021)

50. Baker KS, Suu-Ire R, Barr J, Hayman DTS, Broder CC, Horton DL, Durrant C, Murcia PR, Cunningham AA, Wood JLN. 2014 Viral antibody dynamics in a chiropteran host. *J. Anim. Ecol.* **83**, 415-428. (doi: 10.1111/1365-2656.12153)
51. Van DeVanter DR, Warrener P, Bennett L, Schultz ER, Coulter S, Garber RL, Rose TM. 1996 Detection and analysis of diverse herpesviral species by consensus primer PCR. *J. Clin. Microbiol.* **34**, 1666-1671. (doi: 10.1128/JCM.34.7.1666-1671.1996)
52. Frey BJ, Dueck D. 2007 Clustering by passing messages between data points. *Science* **315**, 972-976. (doi: 10.1126/science.1136800)
53. Dorazio RM, Royle JA. 2005 Estimating size and composition of biological communities by modeling the occurrence of species. *J. Am. Stat. Assoc.* **100**, 389–398. (doi: 10.1198/016214505000000015)
54. Dorazio RM, Royle JA, Soderstrom B, Glimskar A. 2006 Estimating species richness and accumulation by modeling species occurrence and detectability. *Ecology* **87**, 842–854. (doi: 10.1890/0012-9658(2006)87[842:ESRAAB]2.0.CO;2)
55. Schmidt BR, Kéry M, Ursenbacher S, Hyman OJ, Collins JP. 2013 Site occupancy models in the analysis of environmental DNA presence/absence surveys: a case study of an emerging amphibian pathogen. *Methods Ecol. Evol.* **4**, 646–653. (doi: 10.1111/2041-210X.12052)

56. Dopheide A, Tooman LK, Grosser S, Agabiti B, Rhode B, Xie D, Stevens MI, Nelson N, Buckley TR, Drummond AJ, Newcomb RD. 2019 Estimating the biodiversity of terrestrial invertebrates on a forested island using DNA barcodes and metabarcoding data. *Ecol. Appl.* **29**, e01877. (doi: 10.1002/eap.1877)
57. Link WA, Sauer JR. 1996 Extremes in ecology: avoiding the misleading effects of sampling variation in summary analyses. *Ecology* **77**, 1633-1640. (doi: 10.2307/2265557)
58. Zipkin EF, DeWan A, Royle JA. 2009 Impacts of forest fragmentation on species richness: a hierarchical approach to community modelling. *J. Appl. Ecol.* **46**, 815-822. (doi: 10.1111/j.1365-2664.2009.01664.x)
59. Iknayan KJ, Tingley MW, Furnas BJ, Beissinger SR. 2014 Detecting diversity: emerging methods to estimate species diversity. *Trends Ecol. Evol.* **29**, 97-106. (doi: 10.1016/j.tree.2013.10.012)
60. Plummer M. 2003 JAGS: a program for analysis of Bayesian graphical models using Gibbs sampling. In *Proceedings of the Third International Workshop on Distributed Statistical Computing* (eds K Hornik, F Leisch, and A Zeileis) Vienna, Austria: Technische Universitat Wien.

61. Su YS, Yajima M. 2015 R2jags: using R to run ‘JAGS’. Available: <https://cran.r-project.org/web/packages/R2jags/R2jags.pdf>
62. Gelman A, Meng XL, Stern H. 1996 Posterior predictive assessment of model fit via realized discrepancies. *Stat. Sin.* **6**, 733-760.
63. R Core Team. 2017 R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
64. Whittaker RH. 1960 Vegetation of the Siskiyou Mountains, Oregon and California. *Ecol. Monogr.* **30**, 279-338. (doi: 10.2307/1943563)
65. MacArthur R, Recher H, Cody M. 1966 On the relation between habitat selection and species diversity. *Am. Nat.* **100**, 319-332. (doi: 10.1086/282425)
66. Vandegrift KJ, Wale N, Epstein JH. 2011 An ecological and conservation perspective on advances in the applied virology of zoonoses. *Viruses* **3**, 379-397. (doi: 10.3390/v3040379)
67. Geoghegan JL, Homes EC. 2017 Predicting virus emergence amid evolutionary noise. *Open Biol.* **7**, 170189. (10.1098/rsob.170189)

68. Pulliam JRC. 2008 Viral host jumps: moving toward a predictive framework. *EcoHealth* **5**, 80-91. (doi: 10.1007/s10393-007-0149-6)
69. LaDeau SL, Glass GE, Hobbs NT, Latimer A, Ostfeld RS. 2011 Data-model fusion to better understand emerging pathogens and improve infectious disease forecasting. *Ecol. Appl.* **21**, 1443-1460. (doi: 10.1890/09-1409.1)
70. Gandon S, Day T, Metcalf CJE, Grenfall BT. 2016 Forecasting epidemiological and evolutionary dynamics of infectious diseases. *Trends Ecol. Evol.* **31**, 776-788. (doi: 10.1016/j.tree.2016.07.010)
71. Moran KR, Fairchild G, Generous N, Hickman K, Osthus D, Priedhorsky R, Hyman J, Del Valle SY. 2016 Epidemic forecasting is messier than weather forecasting: the role of human behavior and internet data streams in epidemic forecast. *J. Infect. Dis.* **214**, S404-S408. (doi: 10.1093/infdis/jiw375)
72. Dietze MC, Fox A, Beck-Johnson LM, Betancourt JL, Hooten MB, Jarnevich CS, Keitt TH, Kenney MA, Laney CM, Larsen LG, et al. 2018 Iterative near-term ecological forecasting: needs, opportunities, and challenges. *Proc. Natl. Acad. Sci.* **115**, 1424-1432. (doi: 10.1073/pnas.1710231115)
73. Scarpino SV, Petri G. 2019 On the predictability of infectious disease outbreaks. *Nature Comm.* **10**, 898. (doi: 10.1038/s41467-019-08616-0)

74. Dietze MC. 2017 *Ecological Forecasting*, Princeton, NJ: Princeton University Press
75. Suttle CA. 2005 Viruses in the sea. *Nature* **437**, 356-361. (doi: 10.1038/nature04160)
76. De Benedictis P, Schultz-Cherry S, Burnham A, Cattoli G. 2011 Astrovirus infections in humans and animals – molecular biology, genetic diversity, and interspecies transmissions. *Infect. Genet. Evol.* **11**, 1529-1544. (doi: 10.1016/j.meegid.2011.07.024)
77. Lopez-Baucells A, Rocha R, Fernandez-Llamazares A. 2017 When bats go viral: negative framings in virological research imperil bat conservation. *Mammal Rev.* **48**, 62-66. (doi: 10.1111/mam.12110)
78. Streicker DG, Recuenco S, Valderrama W, Benavides JG, Vargas I, Pacheco V, Condori REC, Montgomer J, Rupprecht CE, Rohani P, Altizar S. 2012 Ecological and anthropogenic drivers of rabies exposure in vampire bats: implications for transmission and control. *Proc. R. Soc. B* **279**, 3384-3392. (doi: 10.1098/rspb.2012.0538)
79. Amman BR, Nyakarahuka L, McElroy AK, Dodd KA, Sealy TK, Schuh AJ, Shoemaker TR, Balinandi S, Atimnedi P, Kaboyo W, et al. 2014 Marburgvirus resurgence in Kitaka mine bat population after extermination attempts, Uganda. *Emerg. Infect. Dis.* **20**, 1761-1764. (doi: 10.3201/eid2010.140696)

- 723 80. Olival KJ. 2016 To cull, or not to cull, bat is the question. *EcoHealth* **13**, 6-8. (doi:
724 10.1007/s10393-015-1075-7)
725
- 726 81. Racey PA, Entwistle AC. 2003 Conservation Ecology of Bats. In *Bat Ecology* (eds TH Kunz,
727 and MB Fenton) Chicago, IL: The University of Chicago Press.
728
- 729 82. Jones KE, Purvis A, Gittleman JL. 2003 Biological correlates of extinction risk in bats. *Am.*
730 *Nat.* **161**, 601-614. (doi: 10.1086/368289)
731
- 732 83. Tuttle MD. 1976 Population ecology of the gray bat (*Myotis grisescens*): factors influencing
733 early growth and development. *Ecology* **57**, 587-595. (doi: 10.2307/1936443)
734
- 735 84. MacKenzie DI, Royle JA. 2005 Designing occupancy studies: general advice and allocating
736 survey effort. *J. Appl. Ecol.* **42**, 1105–1114. (doi: 10.1111/j.1365-2664.2005.01098.x)

Table 1. Estimates of hyper-parameter means and their upper and lower 95% credible intervals for the community occupancy model of herpesvirus occurrence in bat hosts. Hyper-parameters represent the average community-wide effect across all viral OTUs, and estimates are presented on the logit scale. Estimates for slope parameters with credible intervals that do not overlap zero are presented in bold.

Response	Hyper-parameter	Estimate	Lower CrI	Upper CrI
Detectability	Intercept	0.840	0.549	1.168
-	Inhibition	-0.065	-0.343	0.210
-	Sub-family	0.067	-0.560	0.715
Occupancy	Intercept	-8.702	-10.725	-6.720
-	Cave	-0.208	-1.911	1.112
-	Sex	0.302	-0.058	0.643
-	Reproductive Status	1.394	0.384	2.354
-	Sex * Reproductive Status	-1.166	-1.912	-0.510
-	Forearm Length	-0.041	-0.130	0.047
-	Ectoparasite Intensity	-0.197	-0.489	0.052

Figure 1: Depiction of molecular methods informing use of hierarchical community-level occupancy modeling. cPCR assays were run twice (a1 and a2). During assay 1, all gel bands were cut (b1), cloned (c1), and sequenced (d1). This resulted in OTU-specific presence-absence matrices that informed the response variable y_{ijk} , presence of OTU i in individual j during assay k , for assay one (e1, green shading). These OTU-specific infections were then condensed to general herpesvirus (i.e. non-OTU-specific) presence-absence to inform the response variable Y_{jk} , presence of herpesvirus in individual j during assay k (e1, purple shading). For assay 2, only those gel bands that did not exist in b1 (e.g. Host 5, b2) were cut, cloned (c2) and sequenced (d2). This informed y_{ijk} (e2, green shading) and Y_{jk} (e2 purple shading) for assay 2. Bands that existed in both assays (e.g. Hosts 1 and 3) were not cut, cloned, or sequenced during assay 2, and therefore OTU identity was not known. Consequently, for assay 2, these data only informed Y_{jk} (e2 purple shading). Both response variables for both assays (f) were used in the modeling framework.

Figure 2: α , β , and γ diversity by sex (black and grey shading), species (color), and species-by-sex groups. Rows represent diversity metrics (β_m signifies the multiplicative model, and β_a signifies the additive model). Error bars indicate posterior uncertainty in metrics calculated using 900 z-matrices. P-values represent significance when comparing empirical metrics to a null distribution using the two-step null generation method (see Appendix). P-values associated with comparisons using two additional null distributions, which do not show substantive differences, are included in Appendix Table 3.

f)

	Host 1	Host 2	Host 3	Host 4	Host 5
OTU A	1, NA	0, 0	0, NA	0, 0	0, 0
OTU B	1, NA	0, 0	0, NA	0, 0	0, 0
OTU C	0, NA	0, 0	1, NA	0, 0	0, 0
OTU D	0, NA	0, 0	0, NA	1, 0	0, 0
OTU E	0, NA	0, 0	0, NA	0, 0	0, 1
Herpes	1, 1	0, 0	1, 1	1, 0	0, 1

Figure 1

767

768 **Figure 2**

769 *Online Appendix for:*
770
771 **Accounting for imperfect detection reveals role of host traits in structuring viral diversity**
772 **of a wild bat community**
773
774 Anna R. Sjodin, Simon J. Anthony, Michael R. Willig, and Morgan W. Tingley
775

Description of cave habitats:

The larger cave (Culebrones Cave) is a structurally complex hot cave, with temperatures reaching 40° C and relative humidity at 100%. It is home to an estimated 300,000 bats representing six species: *Pteronotus quadridens*, *P. portoricensis*, *Mormoops blainvillii*, *Monophyllus redmani*, *Erophylla sezekorni*, and *Brachyphylla cavernarum* [1]. The smaller cave (Larva Cave) is a cold cave that is at ambient temperature and less structurally complex compared to Culebrones Cave. It hosts fewer bats (30-200, ARS personal observation), representing two species, *Artibeus jamaicensis* and *Eptesicus fuscus*. No bat species roosting in the larger cave was found roosting in the smaller cave, and vice versa.

Description of trait measurements:

Bats were weighed to the nearest 0.5 grams using a Pesola scale. Using calipers, forearm length was measured to the nearest 0.01 millimeter from the distal tip of the radius to the distal tip of the olecranon process. Ectoparasite intensity was measured from visual inspection of the number of individual ectoparasites and represented as a categorical variable: low (< 5), medium (5–8), high (9–19), and extreme (>19). A categorical measure of intensity was used instead of abundance because best practices to quantify exact abundance require host anesthesia [2]. Voucher specimens were collected for each ectoparasite morpho-species and classified broadly as flies from the Hippoboscoidea superfamily, mites, and ticks, each of which had multiple morpho-species and variable host ranges (Appendix Table 1). Specimens were deposited in the University of Connecticut's Biodiversity Research and Education Collections. For female bats, pregnancy was determined via gentle palpation of the abdomen, and lactation was determined by milk expression or visible mammary formation. Pregnant or lactating bats were considered to be reproductively active. For male bats, reproductive activity was determined by the presence of descended testes.

Appendix references:

1. Rodríguez-Duran, A. 2009 Bat assemblages in the West Indies: the role of caves. In *Island Bats: Evolution, Ecology, and Conservation* (eds TH Fleming and PA Racey) Chicago, IL: University of Chicago Press.
2. Whitaker Jr JO, Ritzi CM, Dick CW. 2009 Collecting and preserving bat ectoparasites for ecological study. In *Ecological and Behavioral Methods for the Study of Bats* (eds TH Kunz and S Parsons) Baltimore, MD: The Johns Hopkins University Press.

Model code:

The following represents statistical code written in the JAGS modeling language for fitting the multi-species occupancy model of viral community occurrence.

```
batmodel <- function() {  
  # Hyper-parameters for linear parameters
```

```

823     p.assay ~ dunif(0, 1)
824
825     mu.a0 ~ dnorm(0, 0.1)
826     mu.bact ~ dnorm(0, 0.1)
827     mu.b0 ~ dnorm(0, 0.1)
828     mu.cave ~ dnorm(0, 0.1)
829     mu.sex ~ dnorm(0, 0.1)
830     mu.repro ~ dnorm(0, 0.1)
831     mu.sexpro ~ dnorm(0, 0.1)
832     mu.forearm ~ dnorm(0, 0.1)
833     mu.ecto ~ dnorm(0, 0.1)
834
835     tau.a0 ~ dgamma(0.1, 0.1)
836     tau.bact ~ dgamma(0.1, 0.1)
837     tau.b0 ~ dgamma(0.1, 0.1)
838     tau.cave ~ dgamma(0.1, 0.1)
839     tau.sex ~ dgamma(0.1, 0.1)
840     tau.repro ~ dgamma(0.1, 0.1)
841     tau.sexpro ~ dgamma(0.1, 0.1)
842     tau.forearm ~ dgamma(0.1, 0.1)
843     tau.ecto ~ dgamma(0.1, 0.1)
844
845     tau.hostb ~ dgamma(0.1, 0.1)
846     tau.hosta ~ dgamma(0.1, 0.1)
847     df ~ dgamma(0.1, 0.1)
848
849     ## Fixed effect for viral-family impact on detection
850     a.family ~ dnorm(0, 0.1)
851
852     ### Process & Observation model
853     # First, a host-loop to identify whether each host has at least 1 virus
854     detected
855     for(l in 1:n.hosts) {
856         Z[l] <- max(z[l, ])
857         mu.Y[l] <- Z[l] * p.assay
858         for(n in 1:n.visits) {
859             Y[l, n] ~ dbin(mu.Y[l], 1)
860         }
861     }
862
863     # Then, a species-loop to identify whether any given OTU is detected
864     for (i in 1:(n.species)) {
865         b.0[i] ~ dt(mu.b0, tau.b0, df)
866         b.cave[i] ~ dnorm(mu.cave, tau.cave)
867         b.sex[i] ~ dnorm(mu.sex, tau.sex)
868         b.repro[i] ~ dnorm(mu.repro, tau.repro)
869         b.sex.repro[i] ~ dnorm(mu.sexpro, tau.sexpro)
870         b.forearm[i] ~ dnorm(mu.forearm, tau.forearm)
871         b.ecto[i] ~ dnorm(mu.ecto, tau.ecto)
872         a.0[i] ~ dnorm(mu.a0, tau.a0)
873         a.bact[i] ~ dnorm(mu.bact, tau.bact)
874
875         for(m in 1:n.host.spp) {
876             b.host[i, m] ~ dnorm(0, tau.hostb)
877             a.host[i, m] ~ dnorm(0, tau.hosta)
878         }
879
880         # Site process model

```



```

881     for (j in 1:n.hosts) {
882         logit(psi[j, i]) <- b.0[i] + b.host[i, host[j]] +
883         b.cave[i]*cave[j] + b.sex[i] * host.sex[j] + b.repro[i] *
884         repro[j] + b.sex.repro[i] * host.sex[j] * repro[j] +
885         b.forearm[i] * forearm[j] + b.ecto[i] * ecto[j]
886         z[j, i] ~ dbin(psi[j, i], 1)
887
888         # Assay detection model
889         for (k in 1:n.visits){
890             logit(theta[j, i, k]) <- a.0[i] + a.host[i, host[j]]
891             + a.bact[i] * bacteria[j] + a.family * family[i]
892             mu.p[j, i, k] <- theta[j, i, k] * z[j, i]
893             y[j, i, k] ~ dbin(mu.p[j, i, k], 1)
894         }
895     }
896 }
897 }

```

899 **Generation of null distribution:**

901 To create the null distribution, all empirical viral richness values from all sampled bats
902 were placed into a pool (Appendix Figure 1). Richness values from individual bats were then
903 selected at random, and without replacement, from the pool and randomly assigned a host
904 treatment. Random draws continued until simulated treatment groups contained the same number
905 of bats (N) as do empirical treatment groups.

906 This simulation method assumes that the empirical relative abundances of each host
907 treatment group represent the true proportion of individuals of that treatment. We evaluated the
908 robustness of results under this assumption by expanding the methodological approach in two
909 ways. First, we repeated the above methods, but sampled from the regional pool *with*
910 replacement. Second, we adopted a two-part simulation in which step one included random
911 selection of a host species by sex treatment as a “donor”, such that each treatment combination
912 had the same chance of providing hosts to the simulated data set. In step two, we randomly drew
913 an infracommunity from the previously selected donor treatment and placed it into a randomly
914 selected “recipient” treatment. The donor treatments were sampled with replacement, and the
915 two-step simulation was repeated until the number of infracommunities in each recipient
916 treatment equaled empirical numbers.

919 **Appendix Table 1:** Ectoparasite morpho-species and their associated hosts
920

Morpho-Species	Host Species	Infection Site	Number of Infected Hosts
Hippoboscoidea Fly 1	<i>Artibeus jamaicensis</i>	Fur	20
Hippoboscoidea Fly 2	<i>Artibeus jamaicensis</i> , <i>Brachyphylla cavernarum</i> , <i>Eptesicus fuscus</i> , <i>Erophylla sezekorni</i> , <i>Monophyllus redmani</i> , <i>Mormoops blainvillii</i> , <i>Pteronotus portoricensis</i> , <i>P. quadridens</i>	Fur	210
Hippoboscoidea Fly 3	<i>Monophyllus redmani</i> , <i>Pteronotus portoricensis</i> , <i>P. quadridens</i> ,	Fur	202
Mite 1	<i>Artibeus jamaicensis</i> , <i>Monophyllus redmani</i>	Ears	51
Mite 2	<i>Pteronotus quadridens</i>	Wings	133
Mite 3	<i>Artibeus jamaicensis</i> , <i>Brachyphylla cavernarum</i> , <i>Eptesicus fuscus</i> , <i>Erophylla sezekorni</i> , <i>Monophyllus redmani</i> , <i>Mormoops blainvillii</i> , <i>Pteronotus portoricensis</i>	Wings	209
Tick 1	<i>Erophylla sezekorni</i> , <i>Monophyllus redmani</i>	Furred body	31
Tick 2	<i>Pteronotus quadridens</i>	Membranes	3

921
922

Appendix Table 2: Observed vs. predicted OTU prevalence, ranked from most prevalent to least prevalent OTU. Bayesian p-values provide estimates of model fit, with values < 0.05 indicating potential lack of fit. In general, the model showed good fit for most OTUs, with evidence of mild lack-of-fit (under-estimation of prevalence) for the rarest and most common OTUs.

OTU	Observed Prevalence	Mean Predicted Prevalence	Bayesian p-value
21	0.062	0.058	0.059
4	0.058	0.053	0.052
5	0.037	0.034	0.163
18	0.029	0.027	0.193
1	0.028	0.024	0.057
12	0.016	0.014	0.067
9	0.015	0.014	0.407
14	0.013	0.012	0.373
24	0.013	0.012	0.432
27	0.012	0.011	0.351
34	0.012	0.011	0.427
23	0.0083	0.0073	0.382
28	0.0074	0.0067	0.499
13	0.0064	0.0061	0.342
16	0.0064	0.0060	0.408
19	0.0055	0.0052	0.379
20	0.0055	0.0052	0.333
30	0.0055	0.0052	0.350
37	0.0055	0.0050	0.494
11	0.0055	0.0043	0.351
6	0.0037	0.0035	0.272
2	0.0028	0.0025	0.303
7	0.0028	0.0025	0.303
15	0.0028	0.0025	0.321
17	0.0028	0.0026	0.222
22	0.0028	0.0026	0.239
38	0.0028	0.0026	0.246
39	0.0028	0.0026	0.260
25	0.0018	0.0017	0.156
29	0.0018	0.0017	0.169
32	0.0018	0.0017	0.162
40	0.0018	0.0017	0.207
3	0.00092	0.00086	0.110
8	0.00092	0.00083	0.132
10	0.00092	0.00086	0.098
26	0.00092	0.00085	0.108
31	0.00092	0.00087	0.087
33	0.00092	0.00084	0.133
35	0.00092	0.00087	0.087
36	0.00092	0.00088	0.096
41	0.00092	0.00087	0.091

Appendix Table 3: P-values associated with simulation analyses for α , β , and γ diversity. Significant p-values are shown with asterisks (* represents significance at 0.05; ** represents significance at 0.001). β_m signifies the multiplicative model, whereas β_a signifies the additive model. Simulated viral metacommunities were generated using three separate methods (columns), as described in the Appendix.

Diversity effect	Sampling Without Replacement	Sampling With Replacement	Two-Step Sampling Process
α - sex	0.103	0.102	0.091
α - species	<0.001**	<0.001**	<0.001**
α - species*sex	<0.001**	<0.001**	<0.001**
β_m - sex	0.021*	0.027*	0.027*
β_m - species	0.576	0.430	0.595
β_a - sex	0.881	0.884	0.878
β_a - species	1.000	0.999	0.999
γ - sex	0.874	0.870	0.864
γ - species	1.000	1.000	1.000

EMPIRICAL														
Species 1					Species 2					Species 3				
N	Male	S	Female	S	Male	S	Female	S		Male	S	Female	S	N
1	A,B,C,D	4	A,B	2	A,D,E,F	4	A,H	2		A,C,L	3	A,L,M	3	1
2	A,B,C,D	4	A,B	2	E,F	2	A,I,J	3		A,L,M	3	L,M,N	3	2
3	B,C	2	C,D	2	A,E,F,G	4	A,I,J,K	4		B,C	2	A,L,M	3	3
4	C,E	2			E,G	2				A,D	2			4
5										A,D	2			5
Alpha	3		2		3		3			2.4		3		
Beta	1.66666667		2		1.66666667		1.66666667			2.5		1.33333333		
Gamma	5		4		5		5			6		4		
Alpha Analyses		F(sp)			Beta Analyses		F(sp)			Gamma Analyses		F(sp)		
		F(sex)												
		F(sp x sex)												
SIMULATION 1														
Species 1					Species 2					Species 3				
N	Male	S	Female	S	Male	S	Female	S		Male	S	Female	S	N
1	A,B	2	E,F	2	A,E,F,G	4	A,D,E,F	4		A,B,C,D	4	A,B	2	1
2	A,H	2	C,E	2	B,C	2	A,B,C,D	4		C,D	2	E,G	2	2
3	L,M,N	3	A,D	2	A,I,J,K	4	B,C	2		A,I,J	3	A,C,L	3	3
4	A,L,M	3			A,D	2				A,L,M	3			4
5										A,L,M	3			5
Alpha	2.5		2		3		3.33333333			3		2.33333333		
Beta	2.4		2.5		3.33333333		1.8			2.66666667		2.57142857		
Gamma	6		5		10		6			8		6		
Alpha Analyses		F(sp)			Beta Analyses		F(sp)			Gamma Analyses		F(sp)		
		F(sex)												
		F(sp x sex)												
SIMULATION ...														
SIMULATION 1000														
Species 1					Species 2					Species 3				
N	Male	S	Female	S	Male	S	Female	S		Male	S	Female	S	N
1	A,B,C,D	4	E,F	2	A,B	2	C,D	2		B,C	2	A,B	2	1
2	A,E,F,G	4	A,I,J,K	4	E,G	2	A,B,C,D	4		C,E	2	A,D,E,F	4	2
3	A,C,L	3	A,D	2	A,D	2	A,I,J	3		A,H	2	A,L,M	3	3
4	A,L,M	3			L,M,N	3				B,C	2			4
5										A,L,M	3			5
Alpha	3.5		2.66666667		2.25		3			2.2		3		
Beta	2.57142857		2.625		3.55555556		2			3.18181818		2.33333333		
Gamma	9		7		8		6			7		7		
Alpha Analyses		F(sp)			Beta Analyses		F(sp)			Gamma Analyses		F(sp)		
		F(sex)												
		F(sp x sex)												

Appendix Figure 1: Simulation analyses for α , β , and γ diversity are shown in a simplified visualization. Example host species are represented by unique number and color combinations (Species 1 is orange, Species 2 is green, and Species 3 is blue), whereas example host sex is represented by shade of color (light for males versus dark for females). Example viral OTUs are represented by letters (A-M). For each simulation, number of hosts sampled per treatment group (N) remained equal to empirical numbers. Empirical patterns of OTU co-occurrence, and therefore OTU richness per host individual (S), also remained identical to that of the empirical data. First, for 900 iterations, a posterior OTU-host matrix prediction (purple) were used to calculate α , β , and γ diversity for each host treatment group, and appropriate F statistics were calculated. Then, for every simulation (1-1000, yellow), sampled infracommunities were randomly drawn from the pool and placed into a treatment group. F statistics were calculated for

948 each level of analysis, and simulated F statistics were used to create null distributions. Empirical
949 F statistics were then compared to the null distribution.