Title

Bat anthropogenic roosting ecology influences taxonomic and geographic predictions of zoonotic risk.

Authors

Briana A Betke[1,2*], Nicole Gottdenker[3], Lauren Ancel Meyers[1], Daniel J Becker[2]

Affiliations

[1]Department of Integrative Biology, University of Texas at Austin, Austin, TX USA

[2]School of Biological Sciences, University of Oklahoma, Norman, OK USA

[3]Department of Veterinary Pathology, College of Veterinary Medicine, University of Georgia, Athens, GA, USA

*Correspondence to: Briana Betke

Email: briana.a.betke-1@ou.edu

Abstract - 199 words

The ability of wildlife to live in anthropogenic structures is widely observed across many animal species. As proximity to humans is an important risk factor for pathogen transmission, anthropogenic roosting may have important consequences for predicting virus spillover and spillback of viruses. For bats, the influence of roosting in anthropogenic structures on predicting virus hosting ability and diversity is poorly understood. Using machine learning, we analyze novel roosting ecology data to assess the importance of anthropogenic roosting in predicting viral outcomes and evaluate how this trait affects the prediction of undetected but likely host species. The importance of anthropogenic roosting varies across viral outcomes, being most important for virus hosting ability and less so for zoonotic virus hosting ability, viral richness, and the proportion of zoonotic viruses harbored by bats. Across viral outcomes, anthropogenic roosting is less important than human population density but more important than most family, diet, and foraging traits. Models with anthropogenic roosting extended the list of undetected zoonotic host species compared to models excluding this trait. Predicted virus host distributions show distinct spatial patterns between anthropogenic and natural roosting bats. These findings suggest anthropogenic roosting has a non-trivial role in predicting viral outcomes in bats.

Keywords: synanthropic, urban, machine learning, zoonotic, macroecology

Introduction

As emerging infectious disease (EID) outbreaks increase [1,2], promptly detecting zoonotic hosts is an important but challenging task. Sampling efforts typically involve resource-intensive field and laboratory efforts to test potential wildlife hosts for the presence of past or active infection with a zoonotic pathogen. Statistical models can minimize logistical and monetary costs by guiding where sampling efforts should be directed at various taxonomic and spatial scales [3–6]. Further, the resulting predictions can lead to testable hypotheses about mechanisms that drive spillover and spillback events. For example, the prediction of frugivorous bats and primates as likely hosts of ebolaviruses by [7,8] informed further research on ecological and phylogenetic drivers of spillover events, finding evidence from historical data that pteropodid bats may serve as primary hosts and cercopithecid primates as secondary hosts of these viruses in Sub-Saharan Africa [7,8].

The ability to accurately identify potential host species is heavily dependent on the quality and quantity of data used to build each predictive model. Typically, such models use species-level network structure or ecological trait data to classify known hosts from negative or unsampled hosts, often using large, standardized datasets. When new predictor data are provided, it is thus necessary to assess potential changes in model performance to optimize pathogen surveillance. Significant efforts have been made to improve host model performance and predictive ability. Examples include incorporating host competency data (i.e., the ability to transmit a pathogen; [9,10]) to narrow predictions, building models on taxonomic subsets of data to improve predictive capacity [11], and assigning pseudo-negatives based on geographic overlap of known hosts to reduce uncertainty and noise in host–pathogen data [12]. While these efforts

are strides toward improved model prediction and performance, data deficiencies limit our ability

to further improve such models that effectively distinguish hosts from non-host species.

One such data deficiency is information on interfaces where humans and wildlife interact.

The observed increase in EID outbreaks has been attributed to greater contact between humans

and wildlife where they geographically overlap, known as human–wildlife interfaces (HWIs;

[13,14]). Processes such as urbanization increasingly drive some wildlife to live in anthropogenic

structures (e.g., buildings, bridges, homes, and tunnels) in response to rapid habitat loss [15–18].

The propensity of wildlife to live in these anthropogenic structures may have consequences for

pathogen transmission by increasing contact between host species and facilitating introduction to

novel hosts, including humans. Incorporating data on the occupancy of anthropogenic structures

in predictive models could in turn inform sampling efforts by identifying high-priority potential

hosts that are most likely to come into contact with humans and locating opportunities for

bidirectional pathogen transmission.

Prior meta-analyses and comparative studies have used urban presence and absence data

to demonstrate the propensity for urban-attracted species to have greater pathogen diversity

[19–22]. However, this body of research relies on broad dichotomies of urban presence or

human population density as proxies of human–wildlife contact. These proxies do not adequately

represent the physical structures used as HWIs, because they either omit observations of animals

in anthropogenic structures (e.g., [23]) or do not clearly distinguish animals existing in urban

areas for long periods of time from those that may only ephemerally visit urban areas (e.g., [22]).

Additionally, data documenting visitor and dweller status exists [24] but does not

comprehensively cover mammals. Such omissions can impact model prediction by identifying

host species present in urban areas that have low probability of interacting with humans in ways

that facilitate pathogen exposure. For instance, humans can be exposed to pathogens through excess droppings when walking under bridges or in buildings where bats live in high densities [25–27] as opposed to encountering wildlife that visit urban areas but typically avoid humans such as mountain lions [28]. Further, most work focuses on intra- rather than interspecific comparisons, and the coverage of urban presence data at the species-level is poor, challenging our ability to assess synanthropic behavior as a predictor in models. One rare interspecific comparison found that rodents that lived in or near anthropogenic structures increased their probability of being zoonotic hosts, likely owing to factors that relate to their fast life histories (i.e., frequent and large litters, large fluctuations in population size; [29]). Additional investigation is needed to identify general patterns in the relationship between dwelling in anthropogenic structures and pathogen outcomes across taxa to fully assess risk of both pathogen spillover and spillback events.

Bats and their associated viruses are a timely system to assess the epidemiological importance of wildlife that occupy anthropogenic structures in predictive models. Due to their higher mutation rates and transmissibility than other pathogens, viruses are responsible for most EIDs [30]. Many bat species play a role in the transmission of these viral pathogens and have well-documented associations with high-profile zoonotic viruses (e.g., Hendra virus, Nipah virus, Marburg virus, SARS-like coronaviruses). For this reason, ample trait data and viral association data exist for bats due to increased surveillance. Previous studies have applied these data to models with the intention of identifying highly predictive traits of viral diversity and detection of novel hosts. This body of work provides a list of known factors such as geographic range size, body mass and size, fecundity, and biogeography [31–33] to compare against the importance of anthropogenic roosting.

In addition to data availability, bats are sensitive to environmental changes, and their behavioral ecology, particularly roost selection, is changing alongside urbanization with epidemiologically relevant implications. The redistribution and reduction of available roosting structures due to urbanization have led to four in 10 extant species to occupy anthropogenic structures [34]. Generally, anthropogenic roosting can affect viral dynamics through increased crowding, fracturing of roosts, and impairment of bat immune responses [35,36]. For example, Mexican free-tailed bats (*Tadarida brasiliensis*) occupying bridges in Texas showed lower T cell responses than their cave-roosting counterparts [36], possibly owing to smaller colony sizes and ectoparasitism in bridges than caves [37]. Anthropogenic roosting bats are also subject to increased human interactions, resulting in interactions that may further facilitate viral spillover and spillback. For instance, a survey in rural Kenya showed that 38% of humans sharing buildings with bats reported engaging in both direct (i.e., touching, bites, and scratches) and indirect (i.e., urine and feces) contact [38]. Furthermore, forced evacuation of bats from their anthropogenic roosts can cause movement into other unoccupied buildings, facilitating spatial spread of viruses such as rabies virus among wild bat populations [39].

Wildlife-occupied anthropogenic structures represent an increasingly critical frontier for bidirectional pathogen transmission between humans and wildlife. Given this potential for direct contact with humans and rapid urbanization occurring globally, we argue that anthropogenic roosting data should be incorporated into models to improve prediction of hosting ability and viral richness. Here, we draw from a novel roosting ecology dataset (natural vs. anthropogenic) compiled in our previous work to compare the importance of anthropogenic roosting to other factors with known associations with viral outcomes (i.e., viral richness, zoonotic proportion, and hosting ability). We then quantified model performance with and without anthropogenic

roosting to test if predictive ability improves with the addition of this epidemiologically relevant and novel trait. Finally, we tested how such models with and without this trait differed in their predictions by also comparing predicted undetected virus and zoonotic virus host species. We expect that inclusion of this would significantly improve host and viral richness models due to greater sampling effort of anthropogenic roosting species, more frequent spillover between species (including humans) in these interfaces, or impacts on infection dynamics within such species noted above.

Methods

Virus data

Virus associations were extracted from The Global Virome in One Network (VIRION) database (v0.2.1) and matched to the 1,279 extant bat species from the most recent bat phylogeny, representing 86% of the 1,482 recognized species in the other Chiroptera [40–42]. VIRION is a comprehensive database that aggregates several host–virus association datasets to create a dynamic and up-to-date collection of host–virus associations with standardized viral taxonomy. These sources include CLOVER, the USAID Emerging Pandemic Threats PREDICT program, and NCBI GenBank. We restricted VIRION to NCBI-verified host and virus species names. Because serological data may lead to overcounting viral associations given known issues of cross-reactivity [43], we further restricted data to those virus associations detected by PCR and isolation methods. Two measures of virus richness were derived from VIRION: total viral richness and zoonotic virus richness. Total viral richness is the number of all known viral species for a given bat species [22]. Zoonotic viruses were identified as any bat viruses also found in humans within VIRION. To identify the trait profile of bat species that disproportionately host zoonotic viruses specifically, we then derived the proportion of zoonotic viruses per host species.

Two binary variables for virus and zoonotic host status were derived from the richness variables to predict potential host and zoonotic host species. Any bat species without viral associations in VIRION were designated values of zero (i.e., a pseudoabsense).

Anthropogenic roosting data

We used an established dataset of anthropogenic roosting ecology manually compiled from existing literature and historical records [34]. This dataset consists of roosting ecology across 80% ($n = 1,029$) of the 1,279 bat species included in the recent phylogeny and is well represented across bat families and biogeographic regions. Of these 1,029 bat species, 40% roost in anthropogenic structures and 42% exclusively use natural roosts. Roosting in human-made structures is described as a binary classification between species-level use of anthropogenic roosts (e.g., mines, buildings, homes, bridges, tunnels, roofs, and attics) and of exclusively natural roosts (e.g., palm leaves, caves, tree hollows, foliage).

Matching VIRION data to our anthropogenic roosting dataset revealed a wide range of virus richness (0 to 206 virus species). The bat species with the highest total identified viral richness recorded was *Eidolon helvum* at 206. These data contained 341 bat species with known associations with at least one virus, for which 227 of these species were anthropogenic roosting. Similarly, 205 bat species were found to host at least one zoonotic virus, of which 143 species were anthropogenic roosting. We observed distinct patterns in virus richness across bat families and roosting status (Fig. 1A). Descriptive density plots suggested the families Rhinolophidae, Pteropodidae, Mormoopidae, and Miniopteridae had higher identified virus richness among anthropogenic roosting species; however, these trends were less apparent for zoonotic proportions, except for the families Natalidae and Mormoopidae (Fig. 1B).

Additional bat trait data

For additional bat traits, including but not limited to those included in prior predictive models of

bat virus outcomes [31–33,44], we primarily used COMBINE, a recent repository containing 54

mammalian traits aggregated from several commonly used sources [45]. Traits include important

life history variables from PanTHERIA [46] and foraging attributes from EltonTraits [47], both

of which have commonly been used in prior analyses to identify drivers of pathogen diversity

and identify undetected host species [3,4,9,22,48–50]. We restricted COMBINE to non-imputed

trait data, as imputation of missing trait data may not account for strong biases in coverage [51].

Biogeographic areas were expanded to binary variables for species occupancy in each

biogeographic region (i.e., Afrotropical, Australian, Indomaylan, Nearctic, Neotropical,

Oceanian, and Palearctic).

Additional geographic variables (i.e., geographic area, human population density, mean

precipitation, mean monthly actual evapotranspiration rate [AET], mean monthly potential

evapotranspiration rate [PET], minimum and maximum latitude, and mean temperature) were

collected from the PanTHERIA database. To account for the shared evolutionary history of bat

hosts and their impact on viral outcomes, we generated binary variables of all 18 bat families

[31]. This trait matrix was merged and taxonomically reconciled with the host–virus association

dataset to serve as the trait matrix in this analysis.

Assessing trait data coverage and sampling effort

We evaluated all trait variables for variance and coverage prior to downstream modeling. Any

variables found to have no variation or coverage under 30%, indicating mostly missing values,

were removed from the dataset (Fig. S1, Supporting Information). We thus excluded 36 out of 99

variables from our analysis, leading to 63 total variables in the dataset in addition to our primary

anthropogenic roosting variable (Supporting Information).

Sampling is not equal across hosts and could influence modeling viral outcomes, as some

species are studied more due to greater accessibility, proximity to humans, and awareness.

Citation count is commonly used as a proxy for sampling bias (i.e., sampling intensity) in

trait-based analyses [4,9,22,33,52]. To control for sampling bias in our models, we calculated the

number of overall host citations and virus-related host citations for each bat species. A PubMed

search was conducted in R with the *easyPubMed* package [53] to calculate both measures of

sampling effort for all bat species in our dataset [4,9,22].

Boosted regression tree analysis

We built boosted regression tree (BRT) models with the *gbm* package for each of our four

response variables to identify traits predictive of overall viral richness, the proportion of zoonotic

viruses, and both measures of host status (all viruses and only zoonotic viruses). Bernoulli-,

Poisson-, or Gaussian-distributed error structures were used for binary, count, and proportion

data (normalized with the Freeman–Tukey double arcsine transformation via the *metafor*

package), respectively. BRTs are machine learning algorithms that determine variables of

predictive importance based on an ensemble of regression or classification trees, even in the

presence of missing data and non-linear responses, making them ideal for analyzing ecological

data [54]. Trait-based BRTs have also been used to identify unknown hosts of viruses specifically

and are found to outperform other modeling approaches for the study of host–virus associations

more generally [5].

Before fitting BRT models, we partitioned our data between training (80%) and test (20%) datasets using the *resample* package for internal model validation (measured via the area under the receiver operator characteristic curve [AUC] with the *ROCR* package for binary models), across 50 different random data partitions. Models were run with optimal parameters determined via a grid search with 10-fold cross-validation to prevent overfitting (Fig. S2, Supporting Information). Resulting variable importance values and performance metrics were averaged across these 50 model runs. We extracted importance rankings for anthropogenic roosting and used a linear model to compare ranks of this variable across our four virus outcomes, using the Benjamini–Hochberg correction to adjust *p*-values for pairwise comparisons [55]. We used partial dependence plots to visualize the effect of anthropogenic roosting on each viral outcome, after controlling for other features (including citation counts). Additionally, host citation count and virus-related host citation count were modeled with the same explanatory variables as Poisson-distributed responses to determine if viral outcomes are heavily influenced by study effort [3,9,32].

Model performance and prediction

Performance metrics were calculated for the appropriate error distributions of each of our four response variables (i.e., AUC, sensitivity, specificity, pseudo R^2). We tested how these performance measures varied between models with and without the anthropogenic roosting variable with unpaired *t*-tests, accompanied by Cohen's *d* effect sizes and associated *p*-values. The *p*-values were again adjusted with the Benjamini–Hochberg correction. From our models, we predicted viral richness, zoonotic proportion, and the probability of virus hosting per species after correcting for sampling bias, using mean host citation counts and mean virus-related host

citation counts [5,9]. We compared these predictions between models with and without

anthropogenic roosting using Spearman rank correlation coefficients and density plots. We then

used these predictions to identify unknown host species (i.e., bat species currently without virus

associations), where likely overall hosts and zoonotic virus hosts were identified using thresholds

that maximize the sum of sensitivity and specificity calculated with the *PresenceAbsence*

package [56,57]. After thresholding, these predicted host species were compared between models

with anthropogenic roosting included and excluded. We visualized the spatial distributions of

predicted hosts from anthropogenic roosting models if the predictions differed from those of

models without anthropogenic roosting. Predicted host species were mapped by overlaying

shapefiles obtained from the IUCN Red List database of mammal geographic ranges [58]we

divided each cell of the overlapping ranges by total bat richness to generate diversity maps of

predicted hosts, which [58][33].

Results

Importance of anthropogenic roosting in predicting viral outcomes

Anthropogenic roosting had low-to-medium importance among 62 other features when

predicting bat virus outcomes (Figure 2). Our BRTs predicting virus and zoonotic virus host

status showed that anthropogenic roosting status is less important than human population density

but more important than most family, diet, and foraging traits (see Supplementary Material and

Figure S3). Aligning with past work and subsets of the bat–virus network, other traits relating to

geography, life history, and taxonomy were consistently among the most important predictors of

virus outcomes (Figure S4, Supporting information).

Next, we used variable ranks derived from our BRTs to compare relative importance of

anthropogenic roosting across virus outcomes. Anthropogenic roosting ability was ranked

highest for predicting virus host status, with a median ranking of 35 (range: 19–52) out of 63

variables (Figure 2). The second highest median ranking was the proportion of zoonotic viruses

at 38 (range: 31–44). The median rankings for virus richness and zoonotic host status models

were 39 (range: 30–48) and 39 (range: 31–46), respectively. Pairwise comparisons of average

ranking for roosting ecology between viral outcomes further confirmed that this variable was

more important for predicting viral host status than for predicting zoonotic viral hosting, overall

viral richness, and zoonotic proportion (Figure 2, Table S2). Average ranking of roosting ecology

for predicting virus richness did not differ from zoonotic hosting ability or zoonotic proportion.

Additionally, average ranking of this variable for predicting zoonotic host status did not differ

from zoonotic proportion.

Limited relative importance of anthropogenic roosting was also reflected in partial

dependence plots for the effect of anthropogenic roosting on virus outcomes while fixing all

other trait variables and citation counts. The effect of anthropogenic roosting did not

substantially differ from natural roosting across viral richness, proportion of zoonotic viruses,

and zoonotic host status (Figure 3). However, anthropogenic roosting had an observable but still

weak effect for viral host status, where anthropogenic roosting bats were slightly more likely to

host a virus than natural roosting bats (Figure 3).

Model performance and anthropogenic roosting

BRT models containing anthropogenic roosting predicted overall virus richness with moderately

high accuracy (pseudo-$R^2 = 0.51 \pm 0.04$). Similarly, models with our roosting variable

moderately predicted the proportion of zoonotic viruses (pseudo-$R^2 = 0.36 \pm 0.01$). Virus hosts

were distinguished with high accuracy ($\overline{AUC} = 0.92 \pm 0.003$), high specificity ($\overline{Specificity} =$

0.94 ± 0.003), and moderately high sensitivity ($\overline{Sensitivity}$ = 0.66 ± 0.01). Our models

classified zoonotic hosts in a similar fashion to those of virus hosts ($\overline{AUC}$ = 0.93 ± 0.003,

$\overline{Specificity}$ = 0.97 ± 0.002 , $\overline{Sensitivity}$ = 0.63 ± 0.01).

When comparing the performance of these models to those without anthropogenic

roosting, all differences were negligible. Viral richness models without anthropogenic roosting

had equivalent performance (pseudo-R^2 = 0.50 ± 0.04; t = 0.006, p = 1; Cohen's d = 0.001). We

observed this lack of difference in performance for zoonotic proportion models (pseudo-R^2 =

0.36 ± 0.01; t = -0.02, p = 1; Cohen's d = -0.004) and zoonotic host status ($\overline{AUC}$ = 0.93 ± 0.003; t

= -0.02, p = 1; Cohen's d = -0.005). Similarly, sensitivity ($\overline{Sensitivity}$ = 0.63 ± 0.01; t = -0.04, p

= 1; Cohen's d = -0.01) and specificity ($\overline{Specificity}$ = 0.97 ± 0.002; t = 0, p = 1; Cohen's d = 0)

did not differ between models for zoonotic host status. Virus host models without anthropogenic

roosting also did not vary in performance ($\overline{AUC}$ = 0.92 ± 0.003; t = -0.03, p = 1; Cohen's d =

-0.01), specificity ($\overline{Specificity}$ = 0.94 ± 0.002; t = -0.12, p = 1; Cohen's d = -0.02), or sensitivity

($\overline{Sensitivity}$ = 0.66 ± 0.01; t = -0.07, p = 1; Cohen's d = -0.01).

To assess if the trait profiles of viral outcomes were driven by sampling effort, we ran

BRT models predicting both our citation counts using the same trait predictors as all other

models. We found that the included traits poorly predicted both host citation count (pseudo-R^2 =

0.14 ± 0.07) and virus-related host citation count (pseudo-R^2 = -0.06 ± 0.11).

Anthropogenic roosting and model predictions

Predicted viral richness, zoonotic proportion, virus hosting status, and zoonotic virus hosting

status did not vary between models with and without anthropogenic roosting (ρ = 1, p < 0.001 for

all responses; Figure S9). Despite predictions being very similar between models with and

without anthropogenic roosting, the density of model predictions varied by roosting ecology for each viral outcome. We observed greater predicted values for anthropogenic roosting species for non-binary models (i.e., viral richness and zoonotic proportion; Figure 4) and for binary models (viral hosts and zoonotic hosts, Figure 5).

We further compared the predictions of host models with and without anthropogenic roosting. Our models with this trait predicted 111 virus host species, 65 of which were anthropogenic roosting bats. All but one species was predicted by both models. Models without anthropogenic roosting identified one less host than models with anthropogenic roosting (Table S3). In contrast, our zoonotic host models with roosting ecology identified 27 more predicted species than models without this variable, totaling 189 undetected zoonotic hosts. Of the 189 species predicted, 120 were anthropogenic roosting (Table S4). When we stratified these results by bat family, the Vespertilionidae contained the highest proportion of novel virus and zoonotic virus hosts (42/111 and 80/189 species, respectively). The family with the next-highest number of predicted novel virus host species was the Molosidae (22/111 species). For novel zoonotic host species, the Phyllostomidae held the second highest number of zoonotic hosts (40/189 species). The majority of undetected virus hosts were predicted to be in the Afrotropics and Neotropics, with 39 and 31 species out of 111, respectively. The Neotropics and Afrotropics also contained the most undetected zoonotic hosts, with 78 and 45 of the 189 predicted species, respectively. However, these biogeographical trends for undetected zoonotic hosts differ when accounting for bat diversity.

We lastly mapped the distribution of known and novel hosts predicted from the zoonotic virus host model scaled by total bat diversity to observe spatial differences in the density of anthropogenic and natural roosting hosts (Figure 6). Novel anthropogenic roosting hosts largely

had distinct distributions from the novel hosts that exclusively roost in natural structures. This result was most apparent in West Africa, Scandinavia, and the Arabian Peninsula, where hotspots of predicted anthropogenic roosting zoonotic virus hosts are located.

Discussion

As the rate of pathogen emergence increases [1,2], early detection of hosts is an important first step towards better anticipating spillover events. Predictive models using host traits assist with these efforts by providing targeted guidance for surveillance [5,32]. Such models rely on large, standardized trait datasets and should be evaluated when new information becomes available to ensure optimal predictive ability. In this study, we assessed model behavior with the addition of novel predictor data on the interfaces where humans and wildlife interact. Rather than a broad urban-dwelling criterion [22,24] or variables such as human population density within a species' range to approximate human–wildlife contact [3,4,31], we focus on the ability of a species to live within human-made structures to directly assess relationships at the human–wildlife interface. Bat roosting ecology, particularly the structures in which bats roost, likely serve as important points of cross-species transmission, given bats' involvement in the several emerging infectious disease events [59,60]. Overall, our results suggest that this aspect of bat roosting ecology is only moderately important in predicting viral outcomes (i.e., virus richness, zoonotic proportion, host status, and zoonotic host status) but that the inclusion of this trait provides different taxonomic and geographic predictions of zoonotic risk compared to models without this trait.

The ability of bat species to use anthropogenic roosts was not the most important nor the least important predictive variable across viral outcomes; especially for predicting undetected hosts of at least one virus. Prior models using different facets of roosting ecology similarly found

that roost type did not rank highly in predicting viral outcomes (Guy et al. 2020). Although we cannot directly compare our findings to prior work due to differences in roost type categorizations and roosting data coverage, these results may signal a wider trend in the role roosting ecology plays in predicting species-level trends in bat viral diversity. Further inspecting the influence of our roosting ecology variable on predicting viral outcomes showed that anthropogenic roosting bats had a slightly increased probability of hosting at least one virus. Taken with the results of other viral outcomes, our findings suggest that anthropogenic roosting bats have slightly greater probabilities of hosting viruses than natural roosting bats, but do not differ in their viral richness. This result deviates from prior work showing that urban-adapted mammal species host greater viral richness [22]. Our variable importance and marginal effects plots do not strongly support anthropogenic roosting as a major driver of zoonotic outcomes in particular, suggesting that anthropogenic dwelling may not play as critical of a role in predicting zoonotic hosts for bats as compared to rodents [29].

The weak patterns in variable importance and marginal effects observed here may be explained by differences in the trait profiles of virus hosts compared to other viral outcomes. With the exception of citation count and other key traits (i.e., geographic range and mean human population density), the top determinants of virus richness differed from those of viral hosting ability. Our BRTs predicting virus hosts identified climate variables (i.e., average AET, PET, and temperature) and adult forearm length to be among the top ten predictors, whereas these variables did not rank highly for our virus richness models.

In our previous work, we showed that anthropogenically roosting bat species are characterized by similar climatic factors and larger forearm size [34]. The small difference in marginal effects between anthropogenic and natural roosting bats for viral hosting may therefore

reflect these climactic and morphological similarities, suggesting that the trait profile of anthropogenic roosting bats somewhat aligns with that of viral hosts more than that of bat species with greater viral richness. Body size was a common variable of importance for predicting anthropogenic roosting bats and viral richness. However, bats with high virus richness were characterized by larger body sizes, whereas anthropogenic roosting bats had smaller body sizes. Additionally, other traits may further explain the weak importance of roosting ecology for predicting viral outcomes. For instance, we identified broad geographic range to be a top and positive predictor across viral outcomes, aligning with prior work on bat–virus associations [31–33,44]. However, these distributions capture less of the intermediate ranges than is characteristic of anthropogenic roosting bats [34]. Large litter size was also an important predictor of zoonotic proportion and ability to host zoonotic viruses, whereas anthropogenically roosting bats are characterized by smaller litter sizes [34].

The inclusion of roosting ecology did not substantially improve model performance across viral outcomes, possibly owing to similar models predicting bat virus outcomes already performing well prior to the inclusion of this trait. We built upon similar methodologies from other models of the viral diversity and host status of bats that also show high predictive ability [5,31–33]. This could additionally explain how predictions of viral outcomes from our models incorporating roosting ecology did not deviate from our models without this trait. However, predictions from virus host models with this trait extended the list of undetected hosts despite the small improvement in performance.

Differences in predicted hosts across virus and zoonotic host models reveal that roosting ecology had a greater impact on zoonotic than broader virus host prediction, despite anthropogenic roosting having a greater marginal effect on virus host status. Zoonotic host

models with roosting ecology predicted 27 more undiscovered hosts, whereas virus host models with roosting ecology only detected one additional species than analogous models without this trait. All bat species predicted by the models with roosting ecology were in agreement with the non-roosting ecology models, such that the inclusion of roosting ecology predicted 27 additional species, half of which were anthropogenic roosting bats. This increase in undetected hosts suggests that exclusion of roosting ecology in predictive models can lead to false negatives in terms of detecting zoonotic hosts. Accurately identifying likely virus hosts is especially important both for study design, given the financial and logistical costs of sampling [61], and for bat conservation, by minimizing unnecessary sampling of vulnerable species. For the latter point, anthropogenic roosting bats may be especially vulnerable to biodiversity loss and are further threatened by forced evacuation due to their reliance on human-made structures [34].

Most undetected zoonotic viral hosts were concentrated in the Neotropics and Afrotropics. Seventy-eight of the 189 predicted zoonotic hosts reside in the Neotropics. However, this disproportionate number of hosts in the Neotropics is also a function of greater bat diversity in this region, such that we did not identify the Neotropics as a hotspot relative to global patterns in bat species richness. Despite the lack of empirical support for African bats to serve as reservoir hosts of viruses due to strong sampling biases [62], 45 of our 189 undetected species were predicted to host zoonotic viruses in the Afrotropics. This species list (Table S4) could serve as a useful starting point to geographically expand sampling effort and accurately confirm zoonotic hosts in this region while reducing the logistical costs associated with broad sampling. As many of our predicted species in both the Neotropics and Afrotropics have zero citations (31% and 42%, respectively), our models are capable of identifying potential hosts regardless of poor historic sampling effort. Additionally, over half of these undetected zoonotic

hosts were anthropogenic roosting species. Range maps of these novel zoonotic hosts highlight distinct spatial patterns between natural and anthropogenic roosting bats. While novel hosts across roosting ecologies occur in Central and South America, natural roosting bats are most concentrated in coastal Peru and the temperate forests of Chile and Argentina. Similarly, anthropogenic roosting novel hosts are more broadly distributed across the Eastern Hemisphere than natural roosting novel hosts and are especially concentrated in West Africa, Scandinavia, and the Arabian Peninsula. The prediction of so many undetected hosts and clear differences in their geographic distribution according to roosting ecology in largely undersampled regions highlights the need to prioritize sampling effort particularly in the Afrotropics and Palearctics to better understand bat–virus interactions.

While all host and virus species included in our analyses were verified through the NCBI taxonomy, we do not distinguish between PCR- and isolation-based detections when calculating viral outcomes from the VIRION dataset. Defining host–virus associations more strictly based on host competence, and empirically testing predicted hosts for viral competency, could enhance model predictions and advance our understanding of spillover and spillback risks [9,63]. Furthermore, using models informed by host competence may help mitigate stigma against bats labeled as reservoirs solely based on serological or PCR evidence [62,64].

This study serves as a starting point for understanding the impact of anthropogenic roosting on cross-species transmission and provides several directions for future work. To evaluate anthropogenic roosting at the host species scale, we collapsed the bat–virus network, leading to a loss of information relating to viral traits. The inclusion of viral characteristics such as nucleic acid (DNA vs RNA virus), enveloped vs non-enveloped virus, or transmission mode (i.e., vector, environmental, direct contact) could affect the importance of anthropogenic roosting

for predicting associations with particular viruses. Future models could focus explicitly on link prediction to better integrate both host and virus characteristics into analyses of the host–virus network [65,66]. Additionally, our modeling approach with roosting ecology and extensions into link prediction could readily be adapted to analyses of other pathogens. Bat–bacterial associations are especially understudied, as research effort is increasingly focused on determinants of viral outcomes, but can also have important implications for human and bat health [67,68].

In conclusion, we leveraged recently compiled, novel data on roosting ecology to explore how anthropogenic roosting shapes viral outcomes across bats. While roosting ecology was not a highly predictive trait, our results show that predicted zoonotic viral hosts are distinctly distributed based on their roosting ecology. Our findings also support anthropogenic roosts as a meaningful interface where bat–human contact (and therefore cross-species transmission) occurs [38,39]. A better understanding of how anthropogenic roosting influences bat–virus interactions through the application of macroecological tools and further integration with studies of bat–human interactions within ecological, cultural, behavioral and socioeconomic contexts can provide holistic frameworks to understand as well as prevent transmission of bat-borne viruses and guide field studies in ways that minimize invasive sampling of vulnerable wildlife species.

Author Contributions

Conceptualization, D.B., B.B.; Methodology, D.B. and B.B.; Formal Analysis, B.B.; Investigation, B.B. and D.B.; Data Curation, B.B. and D.B.; Writing–Original Draft, B.B. and D.B., Writing –Review & Editing; B.B., D.B., L.M. and N.G.; Supervision, D.B.

Acknowledgements

B.B. and D.J.B. were supported by funding to the Viral Emergence Research Initiative (VERENA) Institute, including the National Science Foundation (DBI 2515340). B.B. was additionally supported through the National Science Foundation Postdoctoral Fellowship in Biology (DBI 2305782).

Declaration of Interests

The authors declare no competing interests.

Data Availability

The viral association dataset used in this analysis can be accessed at the VIRION GitHub repository (https://github.com/viralemergence/virion). All code and data associated with this analysis can be found at the bathaus GitHub repository (https://github.com/viralemergence/bathaus)[69].

Figures

Figure 1. Distribution of anthropogenic roosting and viral outcomes across bat families.

(A) Density ridgeline plots of log virus richness of bats faceted by families with observations greater than 5. (B) Density ridgeline plots of zoonotic virus richness faceted by families with observations greater than 5. Distributions are colored by anthropogenic (green), natural (purple), and unknown roosting bats (gray).

487 **Figure 2. Differences in ranking of our anthropogenic roosting variable across viral**

488 **outcomes.** Colored dots represent rankings for each individual model run (50 random splits of

489 training and test data). Violin plots show the distribution of rankings for each respective viral

490 outcome. Pairwise comparisons between viral outcomes were evaluated with a linear model with

491 posthoc comparison and p-values adjusted for multiple comparisons.

Figure 3. Partial dependence plots for anthropogenic roosting for all viral outcomes.

Colored dots represent marginal effects for each individual model run (100 random splits of training and test data), and the means are indicated with black squares. Dot colors correspond to viral outcome where virus richness is pink (A), zoonotic proportion is orange (B), virus host status is green (C), and zoonotic host is blue (D).

Figure 4. Densities of predicted viral richness and zoonotic proportion of viruses for models with and without anthropogenic roosting as a variable. The predicted values for known and unknown bat species from viral richness models (A) and zoonotic proportion models (B), stratified by roosting status. Natural roosting, anthropogenic roosting, and unknown roost status hosts are indicated by purple, green, and grey, respectively. The x-axes use cubed root transformation for both viral richness and zoonotic proportion models.

Figure 5. Predicted probability of hosting a virus for models with and without

anthropogenic roosting as a variable. The predicted probabilities of hosting of known and

unknown hosts from virus host and zoonotic virus host models, stratified by roosting status.

Natural roosting, anthropogenic roosting, and unknown roost status hosts are indicated by purple,

green, and grey, respectively. All axes are plotted on a square root scale.

512 **Figure 6. Distribution of zoonotic virus host diversity.** The first column shows the distribution

513 of known anthropogenic (top left panel) and natural roosting (bottom left panel) hosts of

514 zoonotic viruses. The second column shows the distribution of undiscovered anthropogenic

515 roosting (top right panel) and natural roosting (bottom right panel) hosts. Maps are based on the

516 IUCN range list of mammal geographical ranges.

References

1. Smith KF, Goldberg M, Rosenthal S, Carlson L, Chen J, Chen C, Ramachandran S. 2014 Global rise in human infectious disease outbreaks. *J. R. Soc. Interface* **11**. (doi:10.1098/RSIF.2014.0950)

2. Meadows AJ, Stephenson N, Madhav NK, Oppenheim B. 2023 Historical trends demonstrate a pattern of increasingly frequent and severe spillover events of high-consequence zoonotic viruses. *BMJ Glob. Health* **8**, e012026.

3. Han BA, Schmidt JP, Bowden SE, Drake JM. 2015 Rodent reservoirs of future zoonotic diseases. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 7039–7044.

4. Olival KJ, Hosseini PR, Zambrana-Torrelio C, Ross N, Bogich TL, Daszak P. 2017 Host and viral traits predict zoonotic spillover from mammals. *Nature* **546**, 646–650.

5. Becker DJ *et al.* 2022 Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *The Lancet Microbe* **3**, e625–e637.

6. Becker DJ, Washburne AD, Faust CL, Mordecai EA, Plowright RK. 2019 The problem of scale in the prediction and management of pathogen spillover. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**, 20190224.

7. Sundaram M, Schmidt JP, Han BA, Drake JM, Stephens PR. 2022 Traits, phylogeny and host cell receptors predict Ebolavirus host status among African mammals. *PLoS Negl. Trop. Dis.* **16**, e0010993.

8. Sundaram M, Dorado M, Akaribo B, Filion A, Han BA, Gottdenker NL, Schmidt JP, Drake JM, Stephens PR. 2024 Fruit–frugivore dependencies are important in *Ebolavirus* outbreaks in Sub-Saharan Africa. *Ecography (Cop.)* (doi:10.1111/ecog.06950)

9. Mull N, Carlson CJ, Forbes KM, Becker DJ. 2022 Virus isolation data improve host predictions for New World rodent orthohantaviruses. *J. Anim. Ecol.* **91**, 1290–1302.

10. Becker DJ, Han BA. 2021 The macroecology and evolution of avian competence for Borrelia burgdorferi. *Glob. Ecol. Biogeogr.* **30**, 710–724.

11. Dallas TA, Becker DJ. 2021 Taxonomic resolution affects host-parasite association model performance. *Parasitology* **148**, 584–590.

12. Tonelli A, Blagrove M, Wardeh M, Marco MD. 2024 A framework to predict zoonotic reservoirs under data uncertainty: a case study on betacoronaviruses. *Research Square*. (doi:10.21203/rs.3.rs-4304994/v1)

13. Hassell JM, Begon M, Ward MJ, Fèvre EM. 2017 Urbanization and Disease Emergence: Dynamics at the Wildlife-Livestock-Human Interface. *Trends Ecol. Evol.* **32**, 55–67.

14. Jones BA *et al.* 2013 Zoonosis emergence linked to agricultural intensification and environmental change. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 8399–8404.

15. Imlay TL, Nickerson D, Horn AG. 2018 Temperature and breeding success for Cliff Swallows (Petrochelidon pyrrhonota) nesting on man-made structures: ecological traps? *Can. J. Zool.* **97**, 429–435.

16. Clark KD. 1994 Managing raccoons, skunks, and opossums in urban settings. In *Proceedings of the Sixteenth Vertebrate Pest Conference*, p. 10.

17. Widdows CD, Downs CT. 2016 Urban roost temperatures of large-spotted-genets: The effect of anthropogenic structures. *J. Therm. Biol.* **57**, 66–71.

18. Adam MD, Hayes JP. 2000 Use of Bridges as Night Roosts by Bats in the Oregon Coast Range. *J. Mammal.* **81**, 402–407.

19. Murray MH, Sánchez CA, Becker DJ, Byers KA, Worsley-Tonks KEL, Craft ME. 2019 City sicker? A meta-analysis of wildlife health and urbanization. *Front. Ecol. Environ.* **17**, 575–583.

20. Gibb R, Redding DW, Chin KQ, Donnelly CA, Blackburn TM, Newbold T, Jones KE. 2020 Zoonotic host diversity increases in human-dominated ecosystems. *Nature* **584**, 398–402.

21. Werner CS, Nunn CL. 2020 Effect of urban habitat use on parasitism in mammals: a meta-analysis. *Proceedings of the Royal Society B: Biological Sciences* **287**. (doi:10.1098/RSPB.2020.0397)

22. Albery GF, Carlson CJ, Cohen LE, Eskew EA, Gibb R, Ryan SJ, Sweeny AR, Becker DJ. 2022 Urban-adapted mammal species have more known pathogens. *Nature Ecology & Evolution* , 794–801.

23. Jung K, Threlfall CG. 2018 Trait-dependent tolerance of bats to urbanization: a global meta-analysis. *Proc. Biol. Sci.* **285**. (doi:10.1098/rspb.2018.1222)

24. Santini L, González-Suárez M, Russo D, Gonzalez-Voyer A, von Hardenberg A, Ancillotto L. 2019 One strategy does not fit all: determinants of urban adaptation in mammals. *Ecol. Lett.* **22**, 365–376.

25. Gerbáčová K, Maliničová L, Kisková J, Maslišová V, Uhrin M, Pristaš P. 2020 The Faecal Microbiome of Building-Dwelling Insectivorous Bats (Myotis myotis and Rhinolophus hipposideros) also Contains Antibiotic-Resistant Bacterial Representatives. *Curr. Microbiol.* **77**, 2333–2344.

26. Mengistu TS, Van Wyk I, Oosthuizen M, Cohen L, Wentzel J. 2022 A One Health approach to investigate bats as a potential source of zoonotic mycoses in selected areas of Mpumalanga province, the Republic of South Africa. *Ethiop. Vet. J.* **26**, 143–157.

27. Pavlik I *et al.* 2021 Nontuberculous mycobacteria prevalence in bats' guano from caves and attics of buildings studied by culture and qPCR examinations. *Microorganisms* **9**, 2236.

28. Suraci JP, Clinchy M, Zanette LY, Wilmers CC. 2019 Fear of humans as apex predators has landscape-scale impacts from mountain lions to mice. *Ecol. Lett.* **22**, 1578–1586.

29. Ecke F, Han BA, Hörnfeldt B, Khalil H, Magnusson M, Singh NJ, Ostfeld RS. 2022 Population fluctuations and synanthropy explain transmission risk in rodent-borne zoonoses. *Nat. Commun.* **13**, 7532.

30. Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, Daszak P. 2008 Global trends in emerging infectious diseases. *Nature* **451**, 990–993.

31. Han BA, Schmidt JP, Alexander LW, Bowden SE, Hayman DTS, Drake JM. 2016 Undiscovered Bat Hosts of Filoviruses. *PLoS Negl. Trop. Dis.* **10**, e0004815.

32. Plowright RK, Becker DJ, Crowley DE, Washburne AD, Huang T, Nameer PO, Gurley ES, Han BA. 2019 Prioritizing surveillance of Nipah virus in India. *PLoS Negl. Trop. Dis.* **13**, e0007393.

33. Guy C, Ratcliffe JM, Mideo N. 2020 The influence of bat ecology on viral diversity and reservoir status. *Ecol. Evol.* **10**, 5748–5758.

34. Betke BA, Gottdenker NL, Meyers LA, Becker DJ. 2024 Ecological and evolutionary characteristics of anthropogenic roosting ability in bats of the world. *iScience* **27**, 110369.

35. Plowright RK, Foley P, Field HE, Dobson AP, Foley JE, Eby P, Daszak P. 2011 Urban habituation, ecological connectivity and epidemic dampening: the emergence of Hendra virus from flying foxes (Pteropus spp.). *Proc. Biol. Sci.* **278**, 3703–3712.

36. Allen LC, Turmelle AS, Mendonça MT, Navara KJ, Kunz TH, McCracken GF. 2009 Roosting ecology and variation in adaptive and innate immune system function in the Brazilian free-tailed bat (Tadarida brasiliensis). *J. Comp. Physiol. B* **179**, 315–323.

37. Allen LC, Turmelle AS, Widmaier EP, Hristov NI, McCracken GF, Kunz TH. 2011 Variation in physiological stress between bridge- and cave-roosting Brazilian free-tailed bats: Stress in cave- and bridge-roosting bats. *Conserv. Biol.* **25**, 374–381.

38. Jackson RT, Lunn TJ, DeAnglis IK, Ogola JG, Webala PW, Forbes KM. 2024 Frequent and intense human-bat interactions occur in buildings of rural Kenya. *PLoS Negl. Trop. Dis.* **18**, e0011988.

39. Streicker DG, Franka R, Jackson FR, Rupprecht CE. 2013 Anthropogenic roost switching and rabies virus dynamics in house-roosting big brown bats. *Vector Borne Zoonotic Dis.* **13**, 498–504.

40. Upham NS, Esselstyn JA, Jetz W. 2019 Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biol.* **17**, e3000494.

41. Carlson CJ *et al.* 2022 The Global Virome in One Network (VIRION): an Atlas of Vertebrate-Virus Associations. *MBio* **13**, e0298521.

42. Simmons NB, Cirranello. AL. 2023 Bat Species of the World: A taxonomic and geographic database. Version 1.3.

43. Gilbert AT *et al.* 2013 Deciphering serology to understand the ecology of infectious diseases
in wildlife. *Ecohealth* **10**, 298–313.

44. Turmelle AS, Olival KJ. 2009 Correlates of Viral Richness in Bats (Order Chiroptera).
Ecohealth **6**, 522.

45. Soria CD, Pacifici M, Di Marco M, Stephen SM, Rondinini C. 2021 COMBINE: a
coalesced mammal database of intrinsic and extrinsic traits. *Ecology* **102**, e03344.

46. Jones KE *et al.* 2009 PanTHERIA: a species-level database of life history, ecology, and
geography of extant and recently extinct mammals. *Ecology* **90**, 2648–2648.

47. Wilman H, Belmaker J, Simpson J, de la Rosa C, Rivadeneira MM, Jetz W. 2014 EltonTraits
1.0: Species-level foraging attributes of the world's birds and mammals. *Ecology* **95**,
2027–2027.

48. Luis AD *et al.* 2013 A comparison of bats and rodents as reservoirs of zoonotic viruses: are
bats special? *Proc. Biol. Sci.* **280**, 20122753.

49. Ostfeld RS, Levi T, Jolles AE, Martin LB, Hosseini PR, Keesing F. 2014 Life history and
demographic drivers of reservoir competence for three tick-borne zoonotic pathogens. *PLoS*
One **9**, e107387.

50. Becker DJ, Streicker DG, Altizer S. 2018 Using host species traits to understand the
consequences of resource provisioning for host–parasite interactions. *J. Anim. Ecol.* **87**.
(doi:10.1111/1365-2656.12765)

51. Johnson TF, Isaac NJB, Paviolo A, González-Suárez M. 2021 Handling missing values in
trait data. *Glob. Ecol. Biogeogr.* **30**, 51–62.

52. Guy C, Thiagavel J, Mideo N, Ratcliffe JM. 2019 Phylogeny matters: Revisiting 'a
comparison of bats and rodents as reservoirs of zoonotic viruses'. *Royal Society Open*
Science **6**. (doi:10.1098/rsos.181182)

53. Fantini D. 2019 easyPubMed: Search and Retrieve Scientific Publication Records from
PubMed.

54. Elith J, Leathwick JR, Hastie T. 2008 A working guide to boosted regression trees. *J. Anim.*
Ecol. **77**, 802–813.

55. Benjamini Y, Hochberg Y. 1995 Controlling the false discovery rate: A practical and
powerful approach to multiple testing. *J. R. Stat. Soc.* **57**, 289–300.

56. Liu C, White M, Newell G. 2013 Selecting thresholds for the prediction of species
occurrence with presence-only data. *J. Biogeogr.* **40**, 778–789.

57. Freeman EA, Moisen G. 2008 PresenceAbsence: An R Package for Presence Absence
Analysis. *Journal of Statistical Software*. **23**, 1–31. (doi:10.18637/jss.v023.i11)

58. IUCN. 2022 The IUCN Red List of Threatened Species. Version 2022-2.

59. McKee CD *et al.* 2019 Host Phylogeny, Geographic Overlap, and Roost Sharing Shape
Parasite Communities in European Bats. *Frontiers in Ecology and Evolution* **7**.
(doi:10.3389/fevo.2019.00069)

60. Lunn TJ, Jackson RT, Webala PW, Ogola JG, Forbes KM. 2024 Modern building structures
are a landscape-level driver of bat–human exposure risk in Kenya. *Front. Ecol. Environ.* ,
e2795.

61. Plowright RK, Becker DJ, McCallum H, Manlove KR. 2019 Sampling to elucidate the
dynamics of infections in reservoir hosts. *Philos. Trans. R. Soc. Lond. B Biol. Sci.* **374**,
20180336.

62. Weber N *et al.* 2023 Robust evidence for bats as reservoir hosts is lacking in most African
virus studies: a review and call to optimize sampling and conserve bats. *Biol. Lett.* **19**,
20230358.

63. Becker DJ, Seifert SN, Carlson CJ. 2020 Beyond Infection: Integrating Competence into
Reservoir Host Prediction. *Trends Ecol. Evol.* **0**. (doi:10.1016/j.tree.2020.08.014)

64. Shapiro JT *et al.* 2021 Setting the Terms for Zoonotic Diseases: Effective Communication
for Research, Conservation, and Public Policy. *Viruses* **13**. (doi:10.3390/v13071356)

65. Dallas T, Park AW, Drake JM. 2017 Predicting cryptic links in host-parasite networks. *PLoS*
Comput. Biol. **13**. (doi:10.1371/journal.pcbi.1005557)

66. Albery GF *et al.* 2021 The science of the host–virus network. *Nature Microbiology 2021*
6:12 **6**, 1483–1492.

67. Shaw LP, Wang AD, Dylus D, Meier M, Pogacnik G, Dessimoz C, Balloux F. 2020 The
phylogenetic range of bacterial and viral pathogens of vertebrates. *Mol. Ecol.* **29**,
3361–3379.

68. Szentivanyi T, McKee C, Jones G, Foster JT. 2023 Trends in Bacterial Pathogens of Bats:
Global Distribution and Knowledge Gaps. *Transbound. Emerg. Dis.* **2023**.
(doi:10.1155/2023/9285855)

69. Betke BA, Gottdenker N, Meyers LA, Becker DJ. 2023 The role of anthropogenic roosting
ecology in shaping viral outcomes in bats. *bioRxiv*. , 2023.12.12.571362.
(doi:10.1101/2023.12.12.571362)