Received: 2 September 2021 | Accepted: 16 March 2022

DOI: 10.1111/1365-2656.13694

Journal of Animal Ecology | BRITISH ECOLOGICAL SOCIETY

RESEARCH ARTICLE

Virus isolation data improve host predictions for New World rodent orthohantaviruses

Nathaniel Mull[1] | Colin J. Carlson[2] | Kristian M. Forbes[1] | Daniel J. Becker[3]

[1]Department of Biological Sciences, University of Arkansas, Fayetteville, AR, USA

[2]Center for Global Health Science and Security, Georgetown University Medical Center, Washington, DC, USA

[3]Department of Biology, University of Oklahoma, Norman, OK, USA

Correspondence
Nathaniel Mull
Email: ngmull@uark.edu

Funding information
National Science Foundation, Grant/Award Number: BII 2021909 and DEB 1911925

Handling Editor: Andy Fenton

Abstract

1. Identifying reservoir host species is crucial for understanding the ecology of multi-host pathogens and predicting risks of pathogen spillover from wildlife to people.
2. Predictive models are increasingly used for identifying ecological traits and prioritizing surveillance of likely zoonotic reservoirs, but these often employ different types of evidence for establishing host associations. Comparisons between models with different infection evidence are necessary to guide inferences about the trait profiles of likely hosts and identify which hosts and geographical regions are likely sources of spillover.
3. Here, we use New World rodent–orthohantavirus associations to explore differences in the performance and predictions of models trained on two types of evidence for infection and onward transmission: RT-PCR and live virus isolation data, representing active infections versus host competence, respectively. Orthohantaviruses are primarily carried by muroid rodents and cause the diseases haemorrhagic fever with renal syndrome (HFRS) and hantavirus cardiopulmonary syndrome (HCPS) in humans.
4. We show that although boosted regression tree (BRT) models trained on RT-PCR and live virus isolation data both performed well and capture generally similar trait profiles, rodent phylogeny influenced previously collected RT-PCR data, and BRTs using virus isolation data displayed a narrower list of predicted reservoirs than those using RT-PCR data. BRT models trained on RT-PCR data identified 138 undiscovered hosts and virus isolation models identified 92 undiscovered hosts, with 27 undiscovered hosts identified by both models. Distributions of predicted hosts were concentrated in several different regions for each model, with large discrepancies between evidence types. As a form of validation, virus isolation models independently predicted several orthohantavirus–rodent host associations that had been previously identified through empirical research using RT-PCR.
5. Our model predictions provide a priority list of species and locations for future orthohantavirus sampling. More broadly, these results demonstrate the value of multiple data types for predicting zoonotic pathogen hosts. These methods can be applied across a range of systems to improve our understanding of pathogen maintenance and increase efficiency of pathogen surveillance.

KEYWORDS
boosted regression tree, emerging infectious disease, hantavirus, isolation, RT-PCR, spillover, zoonoses

1 | INTRODUCTION

Most emerging infectious diseases (EIDs) in humans are caused by pathogens that naturally circulate in wildlife and infect multiple host species (Jones et al., 2008; Taylor et al., 2001). Although some zoonotic pathogens are transmissible among recipient hosts (e.g. HIV), most human EID cases are the result of unique spillover events, where humans function as dead-end hosts (e.g. rabies, malaria; Morens et al., 2004). For EIDs, identifying likely reservoir host species (i.e. those that maintain and transmit a particular pathogen; Haydon et al., 2002) is a critical step towards understanding the ecology of multi-host pathogens and predicting risks of cross-species transmission (Plowright et al., 2017; Viana et al., 2014). Statistical models play an important role in this process (Becker et al., 2019; Carlson et al., 2021). For example, ecological trait datasets have facilitated the development of models that can identify the typical phenotypes of reservoir host species, which often display faster life histories (Albery & Becker, 2020; Worsley-Tonks et al., 2020). Characterizing these trait profiles can, in turn, spur development of new hypotheses about the within- and between-host mechanism that facilitate pathogen maintenance (e.g. Han et al., 2015; Han et al., 2020). Additionally, aggregating the distributions of known and predicted reservoir host species through geographical information systems (GIS; Cromley, 2003) can determine regions of especially likely zoonotic spillover risk (Becker & Han, 2021; Han et al., 2016).

Identifying likely reservoir hosts, their ecological characteristics and their distributions can elucidate possible sources of zoonotic exposure. Large-scale surveillance of wildlife, often involving non-targeted sampling of a large diversity and abundance of animals, is commonly conducted shortly after disease outbreaks to search for reservoir hosts (e.g. Leroy et al., 2005; Poon et al., 2005). Such studies are expensive, time-consuming and generally inefficient, particularly when there is little information to direct sampling effort (e.g. Poon et al., 2005; Pourrut et al., 2009; Yob et al., 2001). Therefore, predictive models provide two pragmatic benefits. First, informed predictions provide an efficient means to proactively identify likely reservoir hosts prior to outbreaks and guide surveillance efforts during or following outbreaks (Becker et al., 2022; Plowright et al., 2019). Second, identifying likely reservoirs with models also promotes targeted strategies to prevent or mitigate spillover risk (Sokolow et al., 2019).

Given the importance of statistical models to facilitate identifying likely but undetected reservoir hosts and understanding the ecology of multi-host pathogens, there is a critical need to establish optimum techniques (Becker et al., 2022; Crowley et al., 2020). In particular, significant questions remain about how the level of evidence for infection and ability for onward transmission of pathogens affects model performance and prediction (Becker et al., 2020; Worsley-Tonks et al., 2020). Most predictive models have been developed for viruses and are based on serology data (i.e. virus-specific antibodies), which tend to be abundant due to their relative ease and cost-effectiveness to collect. However, such information often only provides evidence of virus exposure, not necessarily current infection (Gilbert et al., 2013). Polymerase chain reaction (PCR), on the other hand, provides stronger evidence of current infection, and can better predict host competence (i.e. the ability to transmit) than serology data (Tolsá et al., 2018). However, PCR can amplify nonviable virus, and therefore it does not necessarily indicate onward transmission potential (Leland & Ginocchio, 2007). Aside from experimental infections (e.g. Komar et al., 2003), which are rare due to logistical constraints, the gold standard evidence for reservoir host competence is isolation of viable virus (e.g. Corona et al., 2018), which indicates the ability to not only be infected, but also a greater likelihood to shed infectious virus (Leland & Ginocchio, 2007). Our understanding of how these different types of evidence alter model performance and predictions is limited, despite clear differences in establishing host associations, their resulting inference about the ecological traits of dead-end hosts versus reservoirs, and their applied relevance (i.e. for identifying target species in reservoir host searches or interventions).

Orthohantaviruses (*Hantaviridae*, genus *Orthohantavirus*) are an ideal virus group to explore differences in the performance and predictions of models trained on different types of infection evidence, due to their broad implications for human health as zoonotic pathogens, the predicted large number of unidentified viruses (Vaheri et al., 2008) and the varying types of virus infection evidence currently available from wildlife surveys. Additionally, unidentified orthohantavirus host species have not been previously evaluated using predictive models. There are currently 58 described orthohantaviruses, primarily found in rodents (Laenen et al., 2019), many of which cause two main human diseases: haemorrhagic fever with renal syndrome (HFRS, which is common throughout the Old World) and hantavirus cardiopulmonary syndrome (HCPS or HPS, which is common throughout the New World). Because each human case is thought to be an independent spillover event from an infected rodent (Avšič-Županc et al., 2019; Forbes et al., 2018), identifying orthohantavirus reservoir host species is critical for efforts to mitigate human disease.

Most known orthohantaviruses, including all orthohantaviruses that cause disease in humans (Forbes et al., 2018), infect rodents in the families Cricetidae and Muridae (superfamily Muroidea), though several mole- and shrew-borne orthohantaviruses of unknown zoonotic potential have been discovered (Arai et al., 2007; Arai et al., 2008; Kang et al., 2009; Kang et al., 2011). Because

cross-species transmission is generally constrained by phylogenetic distance between host species (Longdon et al., 2014; Streicker et al., 2010), undiscovered orthohantaviruses of human health concern are also likely to be found among muroid rodents. Additionally, although the majority of described North and South American orthohantaviruses cause disease in humans (13/22), knowledge of host relationships is weak for these viruses, and frequent discovery of novel orthohantaviruses indicates a high likelihood of unknown viruses in that part of the world (Mull et al., 2020). Efforts to identify likely but novel orthohantavirus reservoirs would therefore be maximized by focusing on New World muroids. Applying modelling efforts at a fine taxonomic resolution should further improve predictability by reducing statistical noise from the larger mammal phylogeny and life-history traits that are confounded with other host families or orders (Dallas & Becker, 2021).

In this study, we assess how the performance and predictions of statistical models of orthohantavirus associations varies between two types of evidence for the propensity of a muroid rodent species to host orthohantaviruses: (a) reverse transcriptase PCR (RT-PCR), indicating susceptibility to infection but not necessarily ability to generate new infections and (b) live virus isolation, indicating reservoir competence. We note that this definition of reservoir hosts applies strictly to host competence and the capacity to transmit (e.g. Becker et al., 2020; Gervasi et al., 2015; Merrill & Johnson, 2020), in contrast to population-level definitions about pathogen maintenance (e.g. Haydon et al., 2002; Viana et al., 2014). We first characterize phylogenetic signal and taxonomic patterns in orthohantavirus hosting, which can identify clades of species more susceptible to infection or that are truly competent reservoirs. Next, we train machine learning models on muroid phylogenetic and trait data and compare the performance of models applied to both forms of infection evidence to identify undiscovered orthohantavirus hosts. Finally, predicted host distributions were mapped to identify concentrated regions of potential novel hosts and to explore how different evidence types generate distinct landscapes of likely risk, particularly when anthropogenic impacts are considered. Generated results will guide ongoing and future efforts to discover novel orthohantaviruses, their host associations and geographical areas with amplified spillover risk. More broadly, determining effective modelling approaches, specifically the role of different types of data indicating infection and onward transmission to new hosts, is critical to optimize tools for identifying and understanding potential zoonotic threats to human health and security.

2 | MATERIALS AND METHODS

2.1 | Hantavirus data

A systematic literature search was conducted in Web of Science to identify empirical studies that reported orthohantavirus infections in New World muroid rodents via RT-PCR or virus isolation (search queries in Appendix S1; PRISMA diagram in Appendix S2; citations for data used in Appendix S3). We recorded the number of studies per rodent species with each of the following criteria: at least one individual RT-PCR positive; all individuals RT-PCR negative or virus isolation from at least one individual. Because orthohantaviruses cause persistent and chronic infections in rodents (Forbes et al., 2018), serological tests are often used to demonstrate current or recent infection and RT-PCR is performed only on samples from antibody-positive individuals for virus characterization (Vaheri et al., 2008). To preclude false positives in these studies, only rodents that had positive RT-PCR results were considered RT-PCR-positive, and all other individuals were considered RT-PCR negative, even if RT-PCR was not conducted. If a study used only serology without either RT-PCR or live virus isolation attempts, then the study was not included. When studies attempted virus isolation, additional RT-PCR results were recorded for specimen tissue analyses, but not infected cell cultures.

In studies that used archived samples reported in a previous study (for the same level of evidence), those samples were omitted to preclude pseudo-replication; instead, the original study was used. If a subsequent study examined a different level of evidence (e.g. virus isolation vs. RT-PCR), then we treated the two studies as a single report. When the number or description of positive and negative results per species was not clear in an article (including specimens reported at the genus level and outdated taxonomy that now represents multiple species), only definitive results were recorded. We manually matched select rodent species names between our orthohantavirus data and our phylogeny and trait data (see below). Species synonyms are provided in our online data repository. Since several *Rattus* and *Mus* are abundant in the Old and New World, only results derived in the Americas were included. Species without published evidence of orthohantavirus RT-PCR or isolation results were assigned pseudoabsences (Becker et al., 2020). No ethics approval was required for this study.

2.2 | Phylogenetic analyses

We used a recently developed supertree of extant mammals to capture rodent phylogeny (Upham et al., 2019). The tree was simplified to our specified rodent species using the APE package in R (Paradis et al., 2004). Prior to predictive models, we conducted two assessments of phylogenetic signal (i.e. the propensity for related rodent species to be more similar in virus positivity). For both response variables (RT-PCR and virus isolation), we used the CAPER package (Orme et al., 2018) to calculate *D*, where a value of 1 indicates a phylogenetically random trait distribution and a value of 0 indicates phylogenetic clustering under a Brownian motion model of evolution (Fritz & Purvis, 2010). Significant departure from either model was quantified using a randomization test with 1,000 permutations. However, because traits may also arise under a punctuated equilibrium model of evolution, we next used a graph-partitioning algorithm, phylogenetic factorization, to flexibly identify clades

with different propensity to be infected or competent at various taxonomic depths (Washburne et al., 2019). Phylogenetic factorization partitions a given phylogeny by iteratively identifying edges in a tree that maximize an objective function contrasting species separated by the edge. In the case of our Bernoulli-distributed response variables, this objective function is the deviance of a categorical variable indicating clades on either side of each edge in the phylogeny; this categorical variable is the predictor in a series of generalized linear models (Crowley et al., 2020). We performed phylogenetic factorization using the PHYLOFACTOR package, and we determined the number of phylogenetic factors (clades) to retain using a Holm's sequentially rejective 5% cut-off for the family-wise error rate (Holm, 1979).

2.3 | Rodent traits

We used a published dataset of 55 traits describing the morphology, geography, taxonomy and life history of rodent species. Trait data were primarily from PanTHERIA alongside derived covariates including postnatal growth rate, relative age to sexual maturity, relative age at first birth, production and species density (Han et al., 2015; Jones et al., 2009). We also used the PICANTE package to quantify evolutionary distinctiveness, a measure of how isolated a species is within our muroid phylogeny (Kembel et al., 2010; Redding & Mooers, 2006). Finally, we included binary covariates for our muroid rodent genera to represent taxonomy. Given substantial gaps in trait coverage for rodents (Figure S1), we only included predictors with non-zero variance or with data for over 70% of species, resulting in 56 covariates (Table S1). Lastly, we used the EASYPUBMED package (accessed May 2021) to obtain the number of citations per species as a proxy for sampling effort (Fantini, 2019; Olival et al., 2017).

2.4 | Boosted regression trees

We used boosted regression trees (BRTs), a trait-based machine learning algorithm, to classify rodent species as orthohantavirus hosts based on our predictor matrix of traits. BRTs circumvent many statistical issues associated with traditional hypothesis testing (e.g. non-independent data, many predictors, complex interactions, non-randomly missing covariates) and can uncover new and surprising patterns in data to develop testable hypotheses or predictions (Hochachka et al., 2007). A recent comparison among machine learning algorithms, based on predicting likely bat hosts of betacoronaviruses, also demonstrated that trait-based models, and BRTs in particular, vastly outperform network-based models (Becker et al., 2022). Using BRTs, we modelled binomial virus positivity separately for RT-PCR and virus isolation.

BRTs maximize classification accuracy by learning patterns of features that best distinguish positive and negative hosts (Elith et al., 2008). This generates recursive binary splits for randomly sampled predictor variables, and successive trees are built using residuals of the prior best-performing tree as the new response. Boosting generates an ensemble of linked trees, where each achieves increasingly more accurate classification. Prior to analysis, we randomly split data into training (70%) and test (30%) sets, using the RSAMPLE package to perform stratified sampling such that both datasets contained equal proportions of positive labels. Models were then trained with the GBM package (Greenwell et al., 2020), with the maximum number of trees set to 5,000, a learning rate of 0.001 and an interaction depth of three. We used a comprehensive grid search to assess variation in model performance based on alternative hyperparameters, finding that these parameterizations struck an optimal balance between model complexity and multiple measures of BRT performance (Appendix S4; Figure S2; Tables S2–S4). All BRTs used a Bernoulli error distribution and fivefold cross-validation, and we used the ROCR package to quantify accuracy as area under the receiver operator curve (AUC; Sing et al., 2005). We also complemented this measure of model performance by calculating sensitivity and specificity with the INFORMATIONVALUE package (Prabhakaran, 2016). As results can depend on random splits between training and test data, we used 100 stratified partitions to generate an ensemble (Evans et al., 2017), resulting in mean performance measures ($\overline{\text{AUC}}$ for accuracy; $\bar{x}$ for specificity and sensitivity). Lastly, to diagnose if trait profiles of positive species are driven by study effort, we ran a secondary set of BRTs using the same hyperparameters (with the exception of 10,000 total trees) that instead modelled species citation counts as a Poisson response (Plowright et al., 2019).

2.5 | Model performance and prediction

To assess how BRT performance varied between RT-PCR and virus isolation models (Becker et al., 2020), we used a *t*-test to compare each measure of model performance, with *p*-values adjusted using the Benjamini–Hochberg correction (Benjamini & Hochberg, 1995). We also assessed similarity in mean variable importance between models by estimating the Spearman correlation coefficient between feature ranks. Next, we predicted the probability of a species being positive for either response. When predicting species status, we set citation counts per species to their mean across species as a post-hoc method to correct for sampling effort and remove at least some bias (Becker et al., 2022). Lastly, we also estimated the Spearman correlation coefficients for the mean predictions between RT-PCR and virus isolation models.

We used these mean predictions to identify 'false-negative' orthohantavirus hosts (i.e. those without a prior recorded orthohantavirus infection or isolation). We identified taxonomic patterns in predictions using Pagel's λ as an estimate of phylogenetic signal with the CAPER package as well as a secondary phylogenetic factorization to identify clades with different predicted probabilities. To identify potential unknown hosts or competent reservoirs, we estimated a 95% sensitivity threshold using the PRESENCEABSENCE package (Freeman & Moisen, 2008), which can stratify predictions at a 5%

omission rate on known true positives. This threshold, while fairly inclusive, mostly selects species with comparable probabilities of being infected or competent to known hosts.

To visualize the spatial distribution of known and predicted rodent hosts, we used the IUCN Red List database of mammal geographical ranges and overlaid these shapefiles for thresholded species based on RT-PCR and isolation models. These distributions were also mapped against a proxy for cumulative anthropogenic impact on natural systems, given by the SEDAC Last of the Wild database's 2009 Human Footprint map (Venter et al., 2016; Venter et al., 2018). This qualitative descriptor encompasses several geospatial layers that describe anthropogenic impacts with relevance to human exposure to rodents and orthohantaviruses, particularly human occupation (i.e. built up settlements and human population), agricultural intensification (i.e. crop lands and pasture lands) and ecosystem fragmentation (i.e. road and railway density).

3 | RESULTS

3.1 | Phylogenetic patterns

Across our 601 New World muroid rodent species, 9.65% displayed evidence of orthohantavirus infection via RT-PCR, whereas only 2% were found positive for virus isolation (Figure 1). Of the 12 species with virus isolation records, only one (*Microtus pennsylvanicus*) did not have recorded evidence of PCR positivity. We identified intermediate phylogenetic signal in RT-PCR data ($D = 0.83$) but little phylogenetic signal in virus isolation data ($D = 0.90$). For the former, phylogenetic patterns in RT-PCR data departed from both randomness ($p = 0.002$) and Brownian motion ($p < 0.001$), whereas virus isolation data departed from Brownian motion ($p < 0.001$) but not phylogenetic randomness ($p = 0.16$). Results from phylogenetic factorization were qualitatively similar. We identified two rodent clades with greater propensities to have orthohantavirus infections detected via RT-PCR. The whole genus *Oligoryzomys* ($n = 20$) and a subclade of the genus *Peromyscus* ($n = 24$) had 40% and 37.5% of species predicted to be capable of becoming infected, respectively, compared to 9% of the paraphyletic remainder. In contrast, our analyses identified no taxonomic patterns in positive virus isolation results.

3.2 | Model performance

Both infection evidence BRT models distinguished orthohantavirus-positive and orthohantavirus-negative rodent species with high accuracy ($\overline{AUC} = 0.92 \pm 0.003$) and specificity ($\bar{x} = 0.99 \pm 0.001$) but low sensitivity ($\bar{x} = 0.20 \pm 0.01$). However, BRTs trained on virus isolation data performed better ($\overline{AUC} = 0.93 \pm 0.004$) than those

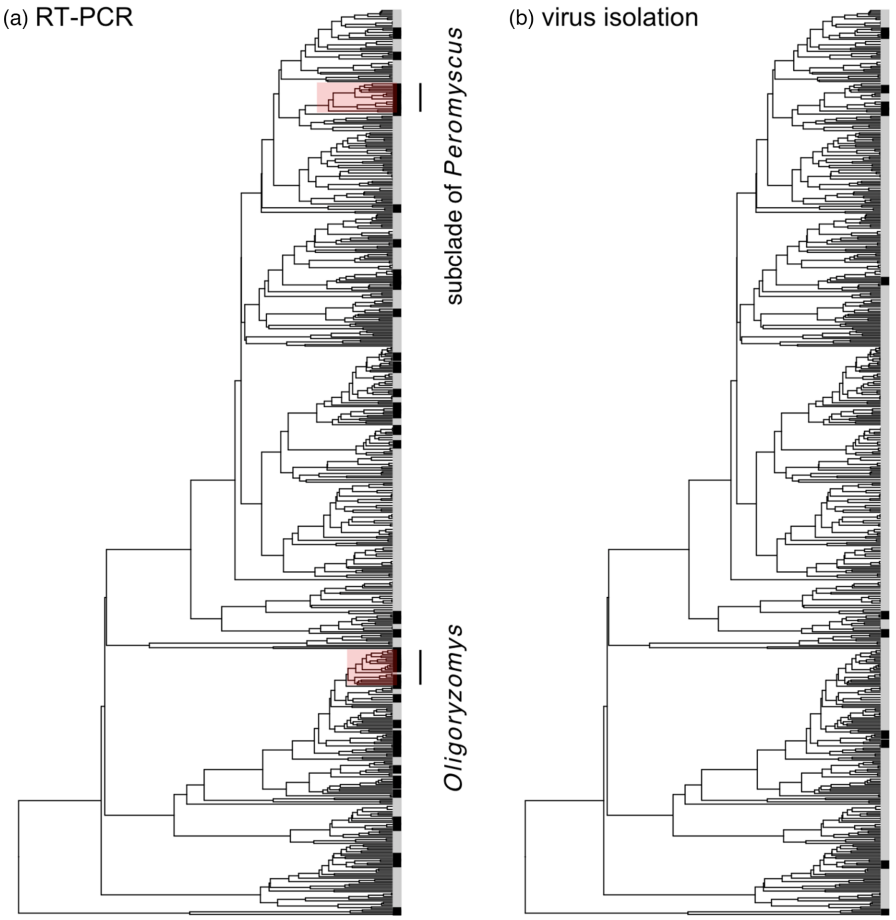

FIGURE 1 Phylogenetic distribution of orthohantavirus-positive muroid rodents in the New World. Species with evidence of infection (a, RT-PCR) or competence (b, live virus isolation) are displayed in black. Visualized in red are any clades identified through phylogenetic factorization for having greater virus positivity when compared to the paraphyletic remainder

trained on RT-PCR data ($\overline{AUC} = 0.91 \pm 0.003$; $t = 2.63$, $p = 0.009$; Figure 2a), resulting in a small standardized effect size (Cohen's $d = 0.37$; Cohen, 1988). RT-PCR models had greater sensitivity ($\bar{x} = 0.37 \pm 0.01$) than virus isolation models ($\bar{x} = 0.03 \pm 0.01$; $t = 22.55$, $p < 0.001$; Cohen's $d = 3.19$; Figure S3), whereas virus isolation models had improved specificity ($\bar{x} = 1.00 \pm 0.00$) over RT-PCR models ($\bar{x} = 0.98 \pm 0.001$; $t = 16.33$, $p < 0.001$; Cohen's $d = 2.31$; Figure S3).

Despite these differences in performance measures, both types of models identified mostly similar species traits as predictive of positivity. Ranks of mean variable importance scores were strongly correlated ($\rho = 0.87$, $p < 0.001$), even after removing traits with zero relative importance ($n = 37$ remaining features; $\rho = 0.84$, $p < 0.001$). Consistently important features for both response variables included PubMed citations, species richness and density within the species range, and evolutionary distinctiveness. Consistently unimportant features included the genera *Thomasomys*, *Rhipidomys*, *Handleyomys* and *Nectomys*. Major discrepancies included the genus *Peromyscus* being an important predictor of RT-PCR positivity but not virus isolation and the genus *Oryzomys* being an important predictor of virus isolation but not RT-PCR positivity (Figure 2b; Table S5). Partial dependence plots suggested that effect directions were largely consistent across models, with positive species being well-studied, located in mammal-rich regions, and characterized by smaller size (Figure S4). Effect direction differed for phylogenetic characteristics, as RT-PCR-positive species were less evolutionarily distinct while species positive for virus isolation were more evolutionarily distinct. Our secondary BRTs of sampling effort showed that citations were not predictable by host traits ($\overline{AUC} = 0.49 \pm 0.001$), suggesting that the trait profiles of positive rodents are not confounded by the traits of well-studied species.

3.3 | Model prediction

Predicted probabilities of being an orthohantavirus host varied widely across the 601 rodent species and were not correlated between BRTs of both evidence types ($\rho = 0.06$, $p < 0.12$; Figure 3a). Many species with intermediate-to-high propensity scores from models based on RT-PCR had a low corresponding probability of being a host based on virus isolation data. Whereas both predictions displayed moderate phylogenetic signal ($\lambda = 0.63$ and 0.57, respectively), the taxonomic patterns identified by phylogenetic factorization largely differed between models (Figure 3b,c; Table S3). For both evidence type models, the genus *Oligoryzomys* ($n = 20$) had a greater mean probability of orthohantavirus hosting compared to the paraphyletic remainder. Predictions from infection models also included a subclade of the genera *Peromyscus* ($n = 25$), *Oxymycterus* ($n = 6$), *Calomys* ($n = 8$) and *Rhipidomys* ($n = 13$) as having higher probabilities ($\bar{x} = 0.62$, 0.74, 0.68 and 0.61, respectively), for which *Peromyscus* was also identified in our phylogenetic factorization of the raw data (Figure 1). However, the subfamily Arvicolinae (including voles, lemmings and muskrats; $n = 43$) had lower probabilities of positive RT-PCR results ($\bar{x} = 0.23$). Predictions from virus isolation models differed from RT-PCR model predictions, as a subclade of the genus *Oecomys* ($n = 7$) and the genus *Oryzomys* ($n = 6$) had greater probabilities of being likely competent reservoirs (both $\bar{x} = 0.24$).

Lastly, we stratified results into binary predictions using a 95% sensitivity threshold. This revealed a total 138 likely undiscovered hosts based on RT-PCR models versus 92 undiscovered hosts based on virus isolation models, of which 27 were also predicted by the former (Table 1). Mapping the geographical distribution of undetected hosts alongside known orthohantavirus-positive rodent species revealed that while predictions from RT-PCR models largely recapitulated the distributions of known RT-PCR-positive species, virus isolation models indicated novel hotspots of overlapping competent reservoirs in the northeastern United States and northern South America, particularly along the Andes Mountains (Figure 4).

4 | DISCUSSION

We used rodent–orthohantavirus associations to demonstrate how statistical model performance and predictions are impacted by different types of infection evidence (i.e. RT-PCR versus virus isolation) alongside identifying rodent species that are likely novel orthohantavirus hosts. Determining the reservoir host of many viruses can be challenging, given logistical challenges of experimental infections and the dramatic variation of infection prevalence across space and time (e.g. for virus isolation; Walsh et al., 2007; Vadell et al., 2011; Holsomback et al., 2013). This is especially true for orthohantaviruses, which are notoriously difficult to isolate (Strandin et al., 2020). However, predictive modelling enables the identification of novel hosts in the absence of field data and, in turn, facilitates targeted field surveillance that can ultimately be used to mitigate hazards posed by zoonotic viruses (Becker et al., 2019). We illustrate here how such models trained on two distinct forms of evidence on host capacity can vary in their performance and predictions regarding likely host species and where they overlap in space.

Orthohantaviruses have traditionally been considered to follow evolutionary cophylogenies with their hosts, with few cross-species infections denoting distinct lineages (Herbreteau et al., 2006; Hjelle et al., 1995; Song et al., 2007). However, the discovery of additional orthohantaviruses has since expanded the diversity of hosts and demonstrated host switches in their evolutionary history (Blasdell et al., 2011; Guo et al., 2013). Indeed, orthohantaviruses have been isolated from species among all four subfamilies of muroid rodents in the Americas. Within those subfamilies, orthohantaviruses have been isolated from seven genera, and the subset of hosts predicted by both models would expand this range by four additional genera (Table 1). In particular, the discovery of an orthohantavirus hosted by *Myodes gapperi* would bridge not only a phylogenetic gap between Eurasian and American viruses, but also a geographical gap between Russia and North America. Several *Microtus* species are the only arvicoline rodents currently known to host orthohantaviruses in the New World, despite a variety of arvicoline hosts in the Old World (Blasdell et al., 2011).

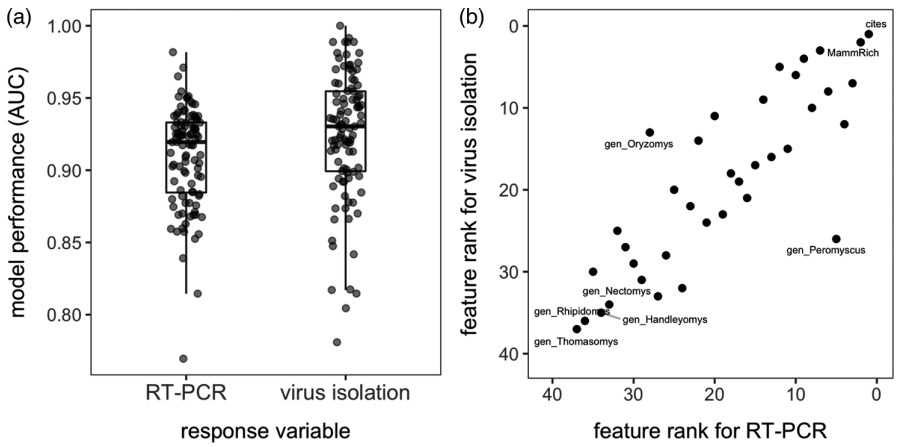

FIGURE 2 Performance of rodent orthohantavirus BRT models trained on RT-PCR versus virus isolation data as the response. (a) Area under the receiver operating characteristic curve (AUC) across 100 random splits of training (70%) and test (30%) data. Boxplots show the median and interquartile range alongside AUC values. (b) Correlation between ranks of mean feature importance between models. Mean relative importance is given in Table S5

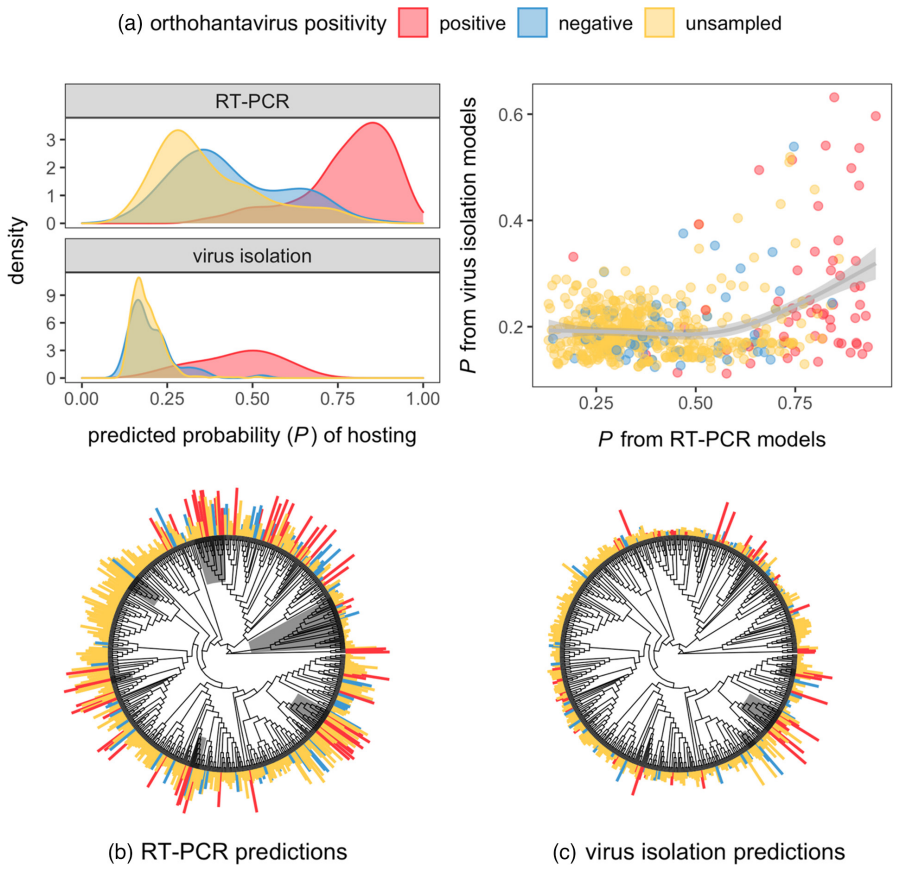

FIGURE 3 Predicted probabilities of rodent orthohantavirus positivity based on RT-PCR versus virus isolation. (a) Distribution of propensity scores stratified by known positive, currently negative and unsampled species. The scatterplot between predictions includes a smoothed curve and confidence intervals from a generalized additive model. (b, c) Taxonomic patterns in predictions as identified through phylogenetic factorization. Segments are scaled by probabilities and coloured as in panel a. Clades identified with significantly different mean predictions are shown in grey, and additional information (e.g. included taxa, species richness) is included in Table S6

In our study, postulated orthohantavirus hosts are mostly concentrated in several regions—southern Mexico and eastern Brazil (based on RT-PCR data) and central and southeastern United States as well as the portions of Peru and northern South America surrounding the Amazon basin (based on virus isolation data; Figure 4). Interestingly, all of these regions coincide with geographical gaps in known orthohantavirus distribution (Guzmán et al., 2017). Distributions from our virus isolation models are consistent with the North American, but not South American, regions predicted to contain rodent reservoirs of novel pathogens at a broader taxonomic scale (Han et al., 2015), and vice versa for distributions from our RT-PCR models, though with a lesser intensity. Such differences highlight the importance of considering various types of evidence, as the use of only one type would have presented an incomplete picture. Future surveillance efforts in these areas will clarify model accuracy to determine the effectiveness of each data type in host predictions.

BRT models trained on RT-PCR and virus isolation data produced similarly high AUC and specificity but did differ in sensitivity. The extremely low sensitivity for virus isolation models is most likely a function of the low predicted probability of hosting based on this form of infection evidence; known positive species had predictions below approximately 60%, which was thus the threshold for classification in calculating sensitivity and specificity. Over-sampling would have likely increased sensitivity (Chawla et al., 2002), but doing so may result in lower specificity (Fountain-Jones et al., 2019). Because the low sensitivity of our models was

TABLE 1 Predicted undiscovered hosts of orthohantaviruses in the Americas: A priority list for future sampling. Plain text species are predicted by the RT-PCR model only, bolded species are predicted by the virus isolation model and starred species are predicted by both models

Genus	Species
Abrothrix	**olivaceus***, **lanosus**, **manni**, **markhami**, **sanborni**
Aepeomys	reigi
Akodon	**boliviensis**, **cursor**, **dayi**, **fumeus**, **lutescens**, iniscatus, leucolimnaeus, mystax, **orophilus**, paranaensis, pervalens, reigi, **subfuscus**, **toba**
Amphinectomys	savamis
Baiomys	musculus, taylori
Brucepattersonius	albinasus, griserufescens, igniventrus, paradisus, soricinus
Calomys	cerqueirai, expulsus, **hummelincki**, musculinus, **sorellus**, **tener**, tocantinsi, venustus
Chibchanomys	orcesi
Chilomys	instans
Delomys	collinus
Dicrostonyx	**groenlandicus**
Eligmodontia	bolsonensis, typus
Euneomys	**chinchilloides**
Euryoryzomys	**macconnelli**, **nitidus**
Graomys	griseoflavus
Habromys	delicatulus, ixtlani, lepturus
Holochilus	brasiliensis, lagigliai, **sciureus**
Hylaeamys	**oniscus**, **yunganus**
Juliomys	rimofrons
Juscelinomys	guaporensis, huanchacae
Lemmus	**trimucronatus**
Loxodontomys	**micropus**
Melanomys	**caliginosus***
Microtus	**chrotorrhinus**, **oeconomus**, **pinetorum**
Microryzomys	**minutus***
Mus	**musculus**
Myodes	**gapperi**, **rutilus**
Neacomys	dubosti, **guianae**, minutus, musseri, paracou, **spinosus***, **tenuipes***
Necromys	**lasiurus**, lenguarum, punctulatus, **urichi***
Nectomys	apicalis, magdalenae, **palmipes**, **rattus**, squamipes
Neotoma	leucodon
Neusticomys	**ferreirai**, **oyapocki**, **peruviensis**, **venezuelae**
Nyctomys	**sumichrasti***
Ochrotomys	**nuttalli**
Oecomys	auyantepui, **bicolor***, catherinae, **concolor***, **mamorae***, paricola, **roberti***, **speciosus**, **superans**, **sydandersoni***, **trinitatis***
Oligoryzomys	**andinus***, arenalis, brendae, **chacoensis**, **delticola***, **destructor***, **eliurus***, **griseolus***, **flavescens**, **magellanicus***, moojeni, **nigripes**, rupestris, **vegetus**, victus
Ondatra	**zibethicus**
Onychomys	torridus
Oryzomys	**antillarum***, **dimidiatus**, **gorgasi**
Otonyctomys	**hatti**
Oxymycterus	amazonicus, angularis, caparoae, dasytrichus, **inca**, josei, paramensis, quaestor, roberti, wayku

(Continues)

TABLE 1 (Continued)

Genus	Species
Peromyscus	carletoni, crinitus, difficilis, fraterculus, **gossypinus**, gratus, guatemalensis, gymnotis, keeni, melanophrys, merriami, mexicanus, nasutus, pembertoni, perfulvus, polionotus, sagax, schmidlyi, simulus
Phenacomys	**ungava**
Phyllotis	**definitus**, **xanthopygus***
Pseudoryzomys	**simplex**
Rattus	**exulans***
Reithrodon	**auritus***
Reithrodontomys	**fulvescens**, **humulis**, **megalotis**, **mexicanus**
Rhagomys	longilingua
Rhipidomys	cariri, caucensis, couesi, emiliae, gardneri, ipukensis, itoan, **leucodactylus***, **macconnelli**, macrurus, mastacalis, modicus, **nitela***, **ochrogaster***, tribei
Sigmodon	fulviventer, **hirsutus***, **inopinatus**, leucotis, planifrons, toltecus, zanjonensis
Synaptomys	**borealis**, **cooperi**
Thomasomys	apeco, **aureus***, caudivarius, cinereiventer, cinereus, cinnameus, eleusis, erro, gracilis, hudsoni, monochromos, onkiro, **oreas**, popayanus, praetor, ucucha, **vestitus**, vulcani
Zygodontomys	**brevicauda**

an artefact of the probability threshold, we opted to prioritize specificity and forego over-sampling. Additionally, the greater AUC for virus isolation models indicated good performance, and species with and without evidence of virus isolation were clearly distinguishable by our BRTs (Figure 3a). In addition to performance, the ecological characteristics identified with positivity through both diagnostic methods were largely similar (Figure 2; Figure S4).

Muroid rodents positive via RT-PCR or virus isolation tended to have smaller body sizes and occur in regions of high mammal richness, alongside other characteristics (Figure S4), matching previous fast pace of life profiles of hosts of zoonotic pathogens across Rodentia more generally (Han et al., 2015). However, in line with our phylogenetic analyses, BRTs also suggested that RT-PCR-positive species were less evolutionarily distinct, likely driven by increased susceptibility to and frequency of spillover events for related species (Longdon et al., 2014; Streicker et al., 2010), whereas virus isolation positive species were more isolated among the muroid phylogeny. Although these differences may be skewed by targeted trapping of known hosts in particular studies (e.g. Safronetz et al., 2008), the broad sampling approach typical of small mammal trapping and hantavirus surveys (e.g. Chu et al., 2008; de Thoisy et al., 2014) suggests that the phylogenetic differences between our RT-PCR and virus isolation models are indicative of the heterogeneity among rodent hosts in ortohantavirus maintenance.

In addition to this phylogenetic contrast, each predictive model generated mostly different lists of potential hosts. There was only minor overlap in predicted host species, and the virus isolation model produced a more concise list of competent reservoir candidates than the RT-PCR model (Table 1). Notably, the isolation model predicted several species that have been identified as orthohantavirus hosts based on empirical studies using RT-PCR (*Holochilus sciureus*, *Loxodontomys micropus*, *Oligoryzomys chacoensis*, *O. flavescens*, *O. nigripes*, *Reithrodontomys fulvescens*, *R. megalotis*, *R. mexicanus* and *Zygodontomys brevicauda*; Mull et al., 2020). These consistencies help validate the predictive capacity of our models, particularly when virus isolation data are included. Ultimately, field studies and natural infection experiments will be necessary to verify our predicted host species, though such model-guided prioritization schemes can provide initial insights to guide empirical efforts (Plowright et al., 2019).

Although this study focused on New World orthohantaviruses to enable higher resolution results in this system, our modelling approach with multiple types of infection evidence is transferable to many other systems. Old World orthohantaviruses represent the most obvious extension, particularly for regions with minimal surveillance, such as Africa, the eastern Mediterranean and Southeast Asia (Guo et al., 2013; Herbreteau et al., 2006). However, other virus groups that pose a threat to human welfare would also benefit from predictive modelling. For example, the reservoir hosts, and likely virus diversity, of orthopoxviruses (e.g. cowpox virus, monkeypox virus) are still mostly unknown, despite common evidence of orthopoxvirus infection among a diverse assemblage of wildlife, particularly rodents (Kinnunen et al., 2011; McInnes et al., 2006) and carnivores (Emerson et al., 2009; Morgan et al., 2019). In such cases, models incorporating multiple levels of infection evidence can help filter out sampling noise caused by spillover to empower host detection for many known and future emerging infectious diseases (Jones et al., 2008). However, this framework is limited to systems with sufficient data from multiple types of evidence, as model training would be challenging and inaccurate, or impossible, for viruses that are rarely or have never been isolated, such as henipaviruses and filoviruses.

Including different levels of infection evidence, and strongly considering data on virus isolation (or other indicators of host

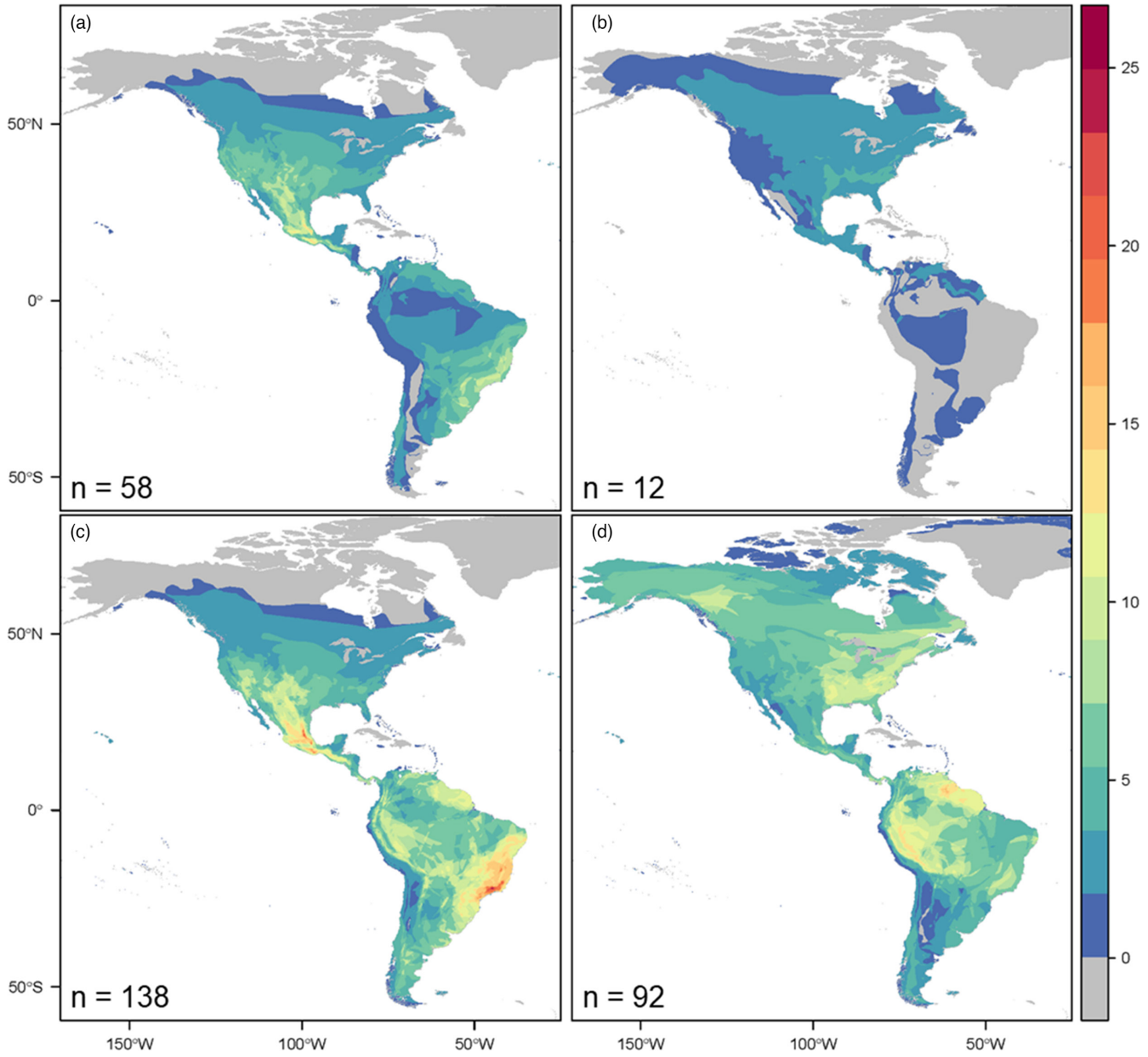

FIGURE 4 Distribution of orthohantavirus hosts. The distribution of known (a, b) and predicted undiscovered (c, d) hosts of orthohantaviruses based on RT-PCR (a, c) and virus isolation (b, d), based on the IUCN Red List database of mammal geographical ranges

competence), will further improve predictive models (Becker et al., 2020). We show here how statistical models trained on two different levels of evidence about infection and the ability to generate new infections both largely performed well and capture mostly similar trait profiles, whereas each model differed most regarding predictions of likely but unsampled host species. Predictions based on viral isolation (i.e. host competence) are most likely to indicate possible reservoir hosts. However, congruent predictions derived from multiple types of evidence indicate particularly notable species and, based on their geographical overlap, regions to consider for future field studies of host ecology, pathogen surveillance and interventions to limit spillover risk. These approaches will improve understanding of pathogen maintenance and increase efficiency in host surveillance not only for orthohantaviruses, but also for many other pathogens important in human and wildlife health.

ACKNOWLEDGEMENTS

This work was supported by funding to the Viral Emergence Research Initiative (VERENA) consortium, including NSF BII 2021909 (C.J.C. and D.J.B.). K.M.F. is supported by NSF DEB 1911925.

CONFLICT OF INTEREST

The authors declare no conflict of interest.

AUTHORS' CONTRIBUTIONS

All authors conceived the initial ideas; N.M. collected the data; D.J.B. and C.J.C. analysed the data; N.M. and D.J.B. wrote the manuscript with input from all authors. All authors contributed to draft revisions and gave final approval for publication.

DATA AVAILABILITY STATEMENT

Data and R code are available from the GitHub repository https://github.com/viralemergence/hantaro and from Zenodo https://doi.org/10.5281/zenodo.6374691 (Mull et al., 2022).

ORCID

Nathaniel Mull https://orcid.org/0000-0002-9888-2229
Colin J. Carlson https://orcid.org/0000-0001-6960-8434
Kristian M. Forbes https://orcid.org/0000-0002-2112-2707
Daniel J. Becker https://orcid.org/0000-0003-4315-8628

REFERENCES

Albery, G. F., & Becker, D. J. (2020). Fast-lived hosts and zoonotic risk. *Trends in Parasitology*, *37*(2), 117–129.

Arai, S., Bennett, S. N., Sumibcay, L., Cook, J. A., Song, J.-W., Hope, A., Parmenter, C., Nerurkar, V. R., Yates, T. L., & Yanagihara, R. (2008). Short report: Phylogenetically distinct hantaviruses in the masked shrew (*Sorex cinereus*) and dusky shrew (*Sorex monticolus*) in the United States. *American Journal of Tropical Medicine and Hygiene*, *78*(2), 348–351.

Arai, S., Song, J.-W., Sumibcay, L., Bennett, S. N., Nerurkar, V. R., Parmenter, C., Cook, J. A., Yates, T. L., & Yanagihara, R. (2007). Hantavirus in northern short-tailed shrew, United States. *Emerging Infectious Diseases*, *13*(9), 1420–1423.

Avšič-Županc, T., Saksia, A., & Korva, M. (2019). Hantavirus infections. *Clinical Microbiology and Infection*, *21*, e6–e16.

Becker, D. J., Albery, G. F., Sjodin, A. R., Poisot, T., Bergner, L. M., Chen, B., Cohen, L. E., Dallas, T. A., Eskew, E. A., Fagre, A. C., Farrell, M. J., Guth, S., Han, B. A., Simmons, N. B., Stock, M., Teeling, E. C., & Carlson, C. J. (2022). Optimising predictive models to prioritise viral discovery in zoonotic reservoirs. *Lancet Microbe*. (in press).

Becker, D. J., & Han, B. A. (2021). The macroecology and evolution of avian competence for *Borrelia burgdorferi*. *Global Ecology and Biogeography*, *30*(3), 710–724.

Becker, D. J., Seifert, S. N., & Carlson, C. J. (2020). Beyond infection: Integrating competence into reservoir host prediction. *Trends in Ecology & Evolution*, *35*(12), 1062–1065.

Becker, D. J., Washburne, A. D., Faust, C. L., Mordecai, E. A., & Plowright, R. K. (2019). The problem of scale in the prediction and management of pathogen spillover. *Philosophical Transactions of the Royal Society B*, *374*(1782), 20190224.

Benjamini, Y., & Hochberg, Y. (1995). Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Methodological)*, *57*(1), 289–300.

Blasdell, K., Henttonen, H., & Buchy, P. (2011). Hantavirus genetic diversity. In S. Morand, F. Beaudeau, & J. Cabaret (Eds.), *New frontiers of molecular epidemiology of infectious diseases* (pp. 179–216). Springer.

Carlson, C., Farrell, M., Grange, Z., Hang, B., Mollentze, N., Phelan, A., Rasmussen, A., Albery, G., Bett, B., Brett-Major, D., Cohen, L., Dallas, T., Eskew, E., Fagre, A., Forbes, K., Gibb, R., Halabi, S., Hammer, C., Katz, R., … Webala, P. W. (2021). Zoonotic risk technology enters the viral emergence toolkit. *Preprints*, 202104.0200. Retrieved from https://www.preprints.org/manuscript/202104.0200/v1

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, *16*, 321–357.

Chu, Y.-K., Owen, R. D., Sánchez-Hernández, C., Romero-Almaraz, M. d. L., & Jonsson, C. B. (2008). Genetic characterization and phylogeny of a hantavirus from western Mexico. *Virus Research*, *131*(2), 180–188.

Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Routledge.

Corona, T. F., Böger, B., da Rocha, T. C., Svoboda, W. K., & Gomes, E. C. (2018). Comparative analysis of mouse inoculation test and virus isolation in cell culture for rabies diagnosis in animals of Parana, Brazil. *Revista da Sociedade Brasileira de Medicina Tropical*, *51*(1), 39–43.

Cromley, E. K. (2003). GIS and disease. *Annual Review of Public Health*, *24*, 7–24.

Crowley, D., Becker, D., Washburne, A., & Plowright, R. (2020). Identifying suspect bat reservoirs of emerging infections. *Vaccine*, *8*(2), 228.

Dallas, T. A., & Becker, D. J. (2021). Taxonomic resolution affects host-parasite association model performance. *Parasitology*, *148*(5), 5845–5890.

de Thoisy, B., Matheus, S., Catzeflis, F., Clément, L., Barrioz, S., Guidez, A., Donato, D., Cornu, J.-F., Brunaux, O., Guitet, S., Lacoste, V., & Lavergne, A. (2014). Short report: Maripa hantavirus in French Guiana: Phylogenetic position and predicted spatial distribution of rodent hosts. *American Journal of Tropical Medicine and Hygiene*, *90*(6), 988–992.

Elith, J., Leathwick, J. R., & Hastie, T. (2008). A working guide to boosted regression trees. *Journal of Animal Ecology*, *77*(4), 802–813.

Emerson, G. L., Yu, L., Frace, M. A., Olsen-Rasmussen, M. A., Khristova, M. L., Govil, D., Sammons, S. A., Regnery, R. L., Karem, K. L., Damon, I. K., & Carroll, D. S. (2009). The phylogenetics and ecology of the orthopoxviruses endemic to North America. *PLoS ONE*, *4*(10), e7666.

Evans, M. V., Dallas, T. A., Han, B. A., Murdock, C. C., & Drake, J. M. (2017). Data-driven identification of potential Zika virus vectors. *eLife*, *6*, 22053.

Fantini, D. (2019). easyPubMed: Search and retrieve scientific publication records from PubMed. R package version 2.13.

Forbes, K. M., Sironen, T., & Plyusnin, A. (2018). Hantavirus maintenance and transmission in reservoir host populations. *Current Opinion in Virology*, *28*, 1–6.

Fountain-Jones, N. M., Machado, G., Carver, S., Packer, C., Recamonde-Mendoza, M., & Craft, M. E. (2019). How to make more from exposure data? An integrated machine learning pipeline to predict pathogen exposure. *Journal of Animal Ecology*, *88*(10), 1447–1461.

Freeman, E. A., & Moisen, G. (2008). PresenceAbsence: An R package for presence absence analysis. *Journal of Statistical Software*, *23*, 31.

Fritz, S. A., & Purvis, A. (2010). Selectivity in mammalian extinction risk and threat types: A new measure of phylogenetic signal strength in binary traits. *Conservation Biology*, *24*(4), 1042–1051.

Gervasi, S. S., Civitello, D. J., Kilvitis, H. J., & Martin, L. B. (2015). The context of host competence: A role for plasticity in host-parasite dynamics. *Trends in Parasitology*, *31*(9), 419–425.

Gilbert, A. T., Jooks, A. R., Haman, D. T. S., Horton, D. L., Müller, T., Plowright, R., Peel, A. J., Bowen, R., Wood, J. L. N., Mills, J., Cunningham, A. A., & Rupprecht, C. E. (2013). Deciphering serology to understand the ecology of infectious diseases in wildlife. *EcoHealth*, *10*(3), 298–313.

Greenwell, B., Boehmke, B., Cunningham, J., & Ridgeway, G.. (2020). gbm: Generalized boosted regression models. R package version 2.1.8.

Guo, W.-P., Lin, X.-D., Wang, W., Tian, J.-H., Cong, M.-L., Zhang, H.-L., Wang, M.-R., Zhou, R.-H., Wang, J.-B., Li, M.-H., Xu, J., Holmes, E.

C., & Zhang, Y.-Z. (2013). Phylogeny and origins of hantaviruses harbored by bats, insectivores, and rodents. *PLoS Pathogens*, *9*(2), e1003159.

Guzmán, C., Calderón, A., González, M., & Mattar, S. (2017). Hantavirus infections. *Revista MVZ Córdoba*, *22*, 6101–6117.

Han, B. A., O'Regan, S. M., Schmidt, J. P., & Drake, J. M. (2020). Integrating data mining and transmission theory in the ecology of infectious diseases. *Ecology Letters*, *23*(8), 1178–1188.

Han, B. A., Schmidt, J. P., Alexander, L. W., Bowden, S. E., Hayman, D. T. S., & Drake, J. M. (2016). Undiscovered bat hosts of filoviruses. *PLoS Neglected Tropical Diseases*, *10*(7), e0004815.

Han, B. A., Schmidt, J. P., Bowden, S. E., & Drake, J. M. (2015). Rodent reservoirs of future zoonotic diseases. *Proceedings of the National Academy of Sciences of the United States of America*, *112*(22), 7039–7044.

Haydon, D. T., Cleveland, S., Taylor, L. H., & Laurenson, M. K. (2002). Identifying reservoirs of infection: A conceptual and practical challenge. *Emerging Infectious Diseases*, *8*(12), 1468–1473.

Herbreteau, V., Gonzalez, J.-P., & Hugot, J.-P. (2006). Implication of phylogenetic systematics of rodent-borne hantaviruses allows understanding of their distribution. *Annals of the New York Academy of Sciences*, *1081*, 39–56.

Hjelle, B., Lee, S.-W., Song, W., Torrez-Martinez, N., Song, J.-W., Yanagihara, R., Gavrilovskaya, I., & Mackow, E. R. (1995). Molecular linkage of hantavirus pulmonary syndrome to the white-footed mouse, *Peromyscus leucopus*: Genetic characterization of the M genome of New York virus. *Journal of Virology*, *69*(12), 8137–8141.

Hochachka, W. M., Caruana, R., Fink, D., Monsun, A. R. T., Riedewald, M., Sorokina, D., & Kelling, S. (2007). Data-mining discovery of pattern and process in ecological systems. *Journal of Wildlife Management*, *71*(7), 2427–2437.

Holm, S. (1979). A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, *6*(2), 65–70.

Holsomback, T. A., Van Nice, C. J., Clark, R. N., McIntyre, N. E., Abuzeineh, A. A., & Salazar-Bravo, J. (2013). Socio-ecology of the marsh rice rat (*Oryzomys palustris*) and the spatio-temporal distribution of Bayou virus in coastal Texas. *Geospatial Health*, *7*(2), 289–298.

Jones, K. E., Bielby, J., Cardillo, M., Fritz, S. A., O'Dell, J., Orme, C. D. L., Safi, K., Sechrest, W., Boakes, E. H., Carbone, C., Connolly, C., Cutts, M. J., Foster, J. K., Grenyer, R., Habib, M., Plaster, C. A., Price, S. A., Rigby, E. A., Rist, J., … Purvis, A. (2009). PanTHERIA: A species-level database of life history, ecology, and geography of extant and recently extinct mammals. *Ecology*, *90*(9), 2648.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature*, *451*(7181), 990–993.

Kang, H. J., Bennett, S. N., Dizney, L., Sumibcay, L., Arai, S., Ruedas, L. A., Song, J.-W., & Yanagihara, R. (2009). Host switch during evolution of a genetically distinct hantavirus in the American shrew mole (*Neurotrichus gibbsii*). *Virology*, *388*(1), 8–14.

Kang, H. J., Bennett, S. N., Hope, A. G., Cook, J. A., & Yanagihara, R. (2011). Shared ancestry between a newfound mole-borne hantavirus and hantaviruses harbored by cricetid rodents. *Journal of Virology*, *85*(15), 7496–7503.

Kembel, S. W., Cowan, P. D., Helmus, M. R., Cornwell, W. K., Morlon, H., Ackerly, D. D., Blomberg, S. P., & Webb, C. O. (2010). Picante: R tools for integrating phylogenies and ecology. *Bioinformatics*, *26*(11), 1463–1464.

Kinnunen, P. M., Henttonen, H., Hoffmann, B., Kallio, E. R., Korthase, C., Laakkonen, J., Niemimaa, J., Palva, A., Schlegel, M., Ali, H. S., Suominen, P., Ulrich, R. G., Vaheri, A., & Vapalahti, O. (2011). Orthopox virus infections in Eurasian wild rodents. *Vector-Borne and Zoonotic Diseases*, *11*(8), 1133–1140.

Komar, N., Langevin, S., Hinten, S., Nemeth, N., Edwards, E., Hettler, D., Davis, B., Bowen, R., & Bunning, M. (2003). Experimental infection of North American birds with the New York 1999 strain of West Nile virus. *Emerging Infectious Diseases*, *9*(3), 311–322.

Laenen, L., Vergote, V., Calisher, C. H., Klempa, B., Klingström, J., Kuhn, J. H., & Maes, P. (2019). *Hantaviridae*: Current classification and future perspectives. *Viruses*, *11*(9), 788.

Leland, D. S., & Ginocchio, C. C. (2007). Role of cell culture for virus detection in the age of technology. *Clinical Microbiology Reviews*, *20*(1), 49–78.

Leroy, E. M., Kumulungui, B., Pourrut, X., Rouquet, P., Hassani, A., Yaba, P., Délicat, A., Paweska, J. T., Gonzalez, J.-P., & Swanepoel, R. (2005). Fruit bats as reservoirs of Ebola virus. *Nature*, *438*(7068), 575–576.

Longdon, B., Brockhurst, M. A., Russell, C. A., Welch, J. J., & Jiggins, F. M. (2014). The evolution and genetics of virus host shifts. *PLoS Pathogens*, *10*(11), e1004395.

McInnes, C. J., Wood, A. R., Thomas, K., Sainsbury, A. W., Gurnell, J., Dein, F. J., & Nettleton, P. F. (2006). Genomic characterization of a novel poxvirus contributing to the decline of the red squirrel (*Sciurus vulgaris*) in the UK. *Journal of General Virology*, *87*(Pt 8), 2115–2125.

Merrill, T. E. S., & Johnson, P. T. J. (2020). Towards a mechanistic understanding of competence: A missing link in diversity-disease research. *Parasitology*, *147*(11), 1159–1170.

Morens, D. M., Folkers, G. K., & Fauci, A. S. (2004). The challenge of emerging and re-emerging infectious diseases. *Nature*, *430*(6996), 242–249.

Morgan, C. N., López-Perez, A. M., Martínez-Duque, P., Jackson, F. R., Suzán, G., & Gallardo-Romero, N. F. (2019). Prevalence of antibodies to orthopoxvirus in wild carnivores of northwestern Chihuahua, Mexico. *Journal of Wildlife Diseases*, *55*(3), 637–644.

Mull, N., Carlson, C. J., Forbes, K. M., & Becker, D. J. (2022). Virus isolation data improve host predictions for New World rodent orthohantaviruses. *Zenodo*, https://doi.org/10.5281/zenodo.6374691

Mull, N., Jackson, R., Sironen, T., & Forbes, K. M. (2020). Ecology of neglected rodent-borne American orthohantaviruses. *Pathogens*, *9*(5), 325.

Olival, K. J., Hosseini, P. R., Zambrana-Torrelio, C., Ross, N., Bogich, T. L., & Daszak, P. (2017). Erratum: Host and viral traits predict zoonotic spillover from mammals. *Nature*, *548*(7669), 612.

Orme, D., Freckleton, R., Thomas, G., Petzoldt, T., Fritz, S., Isaac, N., & Pearse, W. (2018). The Caper package: Comparative analysis of phylogenetics and evolution in R. R package version 1.0.1.

Paradis, E., Claude, J., & Strimmer, K. (2004). APE: Analyses of phylogenetics and evolution in R language. *Bioinformatics*, *20*(2), 289–290.

Plowright, R. K., Becker, D. J., Crowley, D. E., Washburne, A. D., Huang, T., Nameer, P. O., Gurley, E. S., & Han, B. A. (2019). Prioritizing surveillance of Nipah virus in India. *PLoS Neglected Tropical Diseases*, *13*(6), e0007393.

Plowright, R. K., Parrish, C. R., McCallum, H., Hudson, P. J., Ko, A. I., Graham, A. L., & Lloyd-Smith, J. O. (2017). Pathways to zoonotic spillover. *Nature Reviews Microbiology*, *15*(8), 502–510.

Poon, L. L. M., Chu, D. K. W., Chan, K. H., Wong, O. K., Ellis, T. M., Leung, Y. H. C., Lau, S. K. P., Woo, P. C. Y., Suen, K. Y., Yuen, K. Y., Guan, Y., & Peiris, J. S. M. (2005). Identification of a novel coronavirus in bats. *Journal of Virology*, *79*(4), 2001–2009.

Pourrut, X., Souris, M., Towner, J. S., Rollin, P. E., Nichol, S. T., Gonzalez, J.-P., & Leroy, E. (2009). Large serological survey showing circulation of Ebola and Marburg viruses in Gabonese bat populations, and a high seroprevalence of both viruses in *Rousettus aegyptiacus*. *BMC Infectious Diseases*, *9*, 159.

Prabhakaran, S. (2016). InformationValue: Performance analysis and companion functions for binary classification models. R package version 1.2.3.

Redding, D. W., & Mooers, A. Ø. (2006). Incorporating evolutionary measures into conservation prioritization. *Conservation Biology*, *20*(6), 1670–1678.

Safronetz, D., Drebot, M. A., Artsob, H., Cote, T., Makowski, K., & Lindsay, L. R. (2008). Sin Nombre virus shedding patterns in naturally infected deer mice (*Peromyscus maniculatus*) in relation to duration and infection. *Vector-Borne and Zoonotic Diseases*, *8*(1), 97–100.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics*, *21*(20), 3940–3941.

Sokolow, S. H., Nova, N., Pepin, K. M., Peel, A. J., Pulliam, J. R. C., Manlove, K., Cross, P. C., Becker, D. J., Plowright, R. K., McCallum, H., & De Leo, G. A. (2019). Ecological interventions to prevent and manage zoonotic pathogen spillover. *Philosophical Transactions of the Royal Society B*, *374*(1782), 20180342.

Song, J.-W., Baek, L. J., Schmaljohn, C. S., & Yanagihara, R. (2007). Thottapalayam virus, a prototype shrewborne hantavirus. *Emerging Infectious Diseases*, *13*(7), 980–985.

Strandin, T., Smura, T., Ahola, P., Aaltonen, K., Sironen, T., Hepojoki, J., Eckerle, I., Ulrich, R. G., Vapalahti, O., Kipar, A., & Forbes, K. M. (2020). Orthohantavirus isolated in reservoir host cells displays minimal genetic changes and retains wild-type infection properties. *Viruses*, *12*(4), 457.

Streicker, D. G., Trumelle, A. S., Vonhof, M. J., Kuzmin, I. V., McCracken, G. F., & Rupprecht, C. E. (2010). Host phylogeny constrains cross-species emergence and establishment of rabies virus in bats. *Science*, *329*(5992), 676–679.

Taylor, L. H., Latham, S. M., & Woolhouse, M. E. J. (2001). Risk factors for human disease emergence. *Philosophical Transactions of the Royal Society B*, *356*(1411), 983–989.

Tolsá, M. J., García-Peña, G. E., Rico-Chávez, O., Roche, B., & Suzán, G. (2018). Macroecology of birds potentially susceptible to West Nile virus. *Proceedings of the Royal Society B: Biological Sciences*, *285*(1893), 20182178.

Upham, N. S., Esselstyn, J. A., & Jetz, W. (2019). Inferring the mammal tree: Species-level sets of phylogenies for questions in ecology, evolution, and conservation. *PLoS Biology*, *17*(12), e3000494.

Vadell, M. V., Bellomo, C., Martín, A. S., Padula, P., & Villafañe, I. G. (2011). Hantavirus ecology in rodent populations in three protected areas of Argentina. *Tropical Medicine & International Health*, *16*(10), 1342–1352.

Vaheri, A., Vapalahti, O., & Plyusnin, A. (2008). How to diagnose hantavirus infections and detect them in rodents and insectivores. *Reviews in Medical Virology*, *18*(4), 277–288.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. (2018). *Last of the wild project, version 3 (LWP 3): 2009 human footprint, 2018 release*. Socioeconomic Data and Applications Center.

Venter, O., Sanderson, E. W., Magrach, A., Allan, J. R., Beher, J., Jones, K. R., Possingham, H. P., Laurance, W. F., Wood, P., Fekete, B. M., Levy, M. A., & Watson, J. E. M. (2016). Global terrestrial human footpring maps for 1993 and 2009. *Scientific Data*, *3*, 160067.

Viana, M., Mancy, R., Biek, R., Cleveland, S., Cross, P. C., Lloyd-Smith, J. O., & Haydon, D. T. (2014). Assembling evidence for identifying reservoirs of infection. *Trends in Ecology & Evolution*, *29*(5), 270–279.

Walsh, A. S., Louis, T. A., & Glass, G. E. (2007). Detecting multiple levels of effect during survey sampling using a Bayesian approach: Point prevalence estimates of a hantavirus in hispid cotton rats (*Sigmodon hispidus*). *Ecological Modelling*, *205*(1), 29–38.

Washburne, A. D., Silverman, J. D., Morton, J. T., Becker, D. J., Crowley, D., Mukherjee, S., David, L. A., & Plowright, R. K. (2019). Phylofactorization: A graph partitioning algorithm to identify phylogenetic scales of ecological data. *Ecological Monographs*, *89*(2), e01353.

Worsley-Tonks, K. E. L., Escobar, L. E., Biek, R., Castaneda-Guzman, M., Craft, M. E., Streicker, D. G., White, L. A., & Fountain-Jones, N. M. (2020). Using host traits to predict reservoir host species of rabies virus. *PLoS Neglected Tropical Diseases*, *14*(12), e0008940.

Yob, J. M., Field, H., Rashdi, A. M., Morrissy, C., van der Heide, B., Rota, P., bin Adzhar, A., White, J., Daniels, P., Jamaluddin, A., & Ksiazek, T. (2001). Nipah virus infection in bats (order Chiroptera) in Peninsular Malaysia. *Emerging Infectious Diseases*, *7*(3), 439–441.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Mull, N., Carlson, C. J., Forbes, K. M. & Becker, D. J. (2022). Virus isolation data improve host predictions for New World rodent orthohantaviruses. *Journal of Animal Ecology*, *91*, 1290–1302. https://doi.org/10.1111/1365-2656.13694