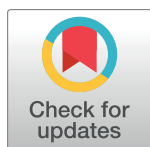RESEARCH ARTICLE

Predicting spatial spread of rabies in skunk populations using surveillance data reported by the public

Kim M. Pepin[1]☯*, Amy J. Davis[1]☯, Daniel G. Streicker[2,3], Justin W. Fischer[1], Kurt C. VerCauteren[1], Amy T. Gilbert[1]

1 National Wildlife Research Center, United States Department of Agriculture, Wildlife Services, Fort Collins, Colorado, United States of America, 2 Institute of Biodiversity, Animal Health and Comparative Medicine, University of Glasgow, Glasgow, Scotland, 3 Medical Research Council-University of Glasgow Centre for Virus Research, Glasgow, Scotland

☯ These authors contributed equally to this work.
* Kim.M.Pepin@aphis.usda.gov

Abstract

Background

Prevention and control of wildlife disease invasions relies on the ability to predict spatio-temporal dynamics and understand the role of factors driving spread rates, such as seasonality and transmission distance. Passive disease surveillance (i.e., case reports by public) is a common method of monitoring emergence of wildlife diseases, but can be challenging to interpret due to spatial biases and limitations in data quantity and quality.

Methodology/Principal findings

We obtained passive rabies surveillance data from dead striped skunks (*Mephitis mephitis*) in an epizootic in northern Colorado, USA. We developed a dynamic patch-occupancy model which predicts spatio-temporal spreading while accounting for heterogeneous sampling. We estimated the distance travelled per transmission event, direction of invasion, rate of spatial spread, and effects of infection density and season. We also estimated mean transmission distance and rates of spatial spread using a phylogeographic approach on a subsample of viral sequences from the same epizootic. Both the occupancy and phylogeographic approaches predicted similar rates of spatio-temporal spread. Estimated mean transmission distances were 2.3 km (95% Highest Posterior Density (HPD$_{95}$): 0.02, 11.9; phylogeographic) and 3.9 km (95% credible intervals (CI$_{95}$): 1.4, 11.3; occupancy). Estimated rates of spatial spread in km/year were: 29.8 (HPD$_{95}$: 20.8, 39.8; phylogeographic, branch velocity, homogenous model), 22.6 (HPD$_{95}$: 15.3, 29.7; phylogeographic, diffusion rate, homogenous model) and 21.1 (CI$_{95}$: 16.7, 25.5; occupancy). Initial colonization probability was twice as high in spring relative to fall.

Conclusions/Significance

Skunk-to-skunk transmission was primarily local (< 4 km) suggesting that if interventions were needed, they could be applied at the wave front. Slower viral invasions of skunk rabies

in western USA compared to a similar epizootic in raccoons in the eastern USA implies host species or landscape factors underlie the dynamics of rabies invasions. Our framework provides a straightforward method for estimating rates of spatial spread of wildlife diseases.

Author summary

Rabies is a deadly zoonotic infection with a global distribution. In 2012, an epizootic of skunk rabies established in northern Colorado, USA and spread rapidly through three counties. The epizootic was documented through reports of dead skunks by the public. We examined the reports to determine how rapidly rabies was moving and which factors could explain the patterns of spread. We compared these estimates of spatial movement of rabies to those obtained from analyzing rabies genetic sequences that we obtained from some of the dead skunks reported by the public. By both methods, we found the virus was moving south at a little over 20 km/year and that most transmission between skunks occurred at short distances (< 4 km). Rabies was most likely to spread to new areas during the first half of the year, when skunk populations were producing new offspring. Our genetic model suggested that roads and rivers in the study landscape did not affect movement speed of rabies. We developed a framework that used the spatial data in the public reports to predict where and when skunk rabies would occur next. This framework could be used on public health surveillance data for other diseases or countries.

Introduction

A central focus for disease ecologists and epidemiologists is to quantify processes that determine geographic spread of disease [1]. Surveillance systems, which rely on reporting by the public [2–5], provide data that can improve understanding disease dynamics and planning interventions. However, passive surveillance data are challenging to interpret because the underlying sampling design is opportunistic. Raw patterns may depend on observation processes that produce a biased representation of disease occurrence. Interpretation of passive surveillance data from wildlife populations can be especially challenging because the underlying ecological processes, such as host population density, distribution, and demographic dynamics, are often unknown [e.g., 6]. In these cases a phenomenological method that does not rely on explicit representation of often unavailable host ecological data, may be valuable for quantifying disease spread—especially when surveillance data are too sparse, or disease prevalence too low, to enable estimation of multiple unknown parameters representing non-linear processes from both the host and disease dynamics.

Rabies virus (RABV) is a globally-distributed zoonotic pathogen that circulates naturally in a variety of carnivore and bat host species and has among the highest case fatality rate of known infectious diseases [7]. The principal burden of human and animal cases is associated with domestic dog populations [8], but emergence in wild carnivores has been observed, especially in areas where domestic dog rabies has been managed or controlled [9,10], or in areas with a long history of disease absence [11]. Risk of rabies transmission from wildlife is traditionally monitored through public health surveillance systems, which involve voluntary reports by the public of domestic animal or human exposures to potentially sick wildlife (appearance of atypical behavior, and especially signs of neurologic illness), followed by diagnostic testing [7]. Animal movement can have significant consequences on geographic spread

of RABV [12], but the transmission distance during individual infections, an important component of geographic spreading, is poorly documented. Analysis of spatial public-health surveillance data have the potential to improve our understanding of the geographic spreading processes of RABVs, which will in turn help with planning prevention and response strategies, and prioritizing resources in space and time.

In the United States of America (USA), rabies infections of humans and animals became nationally reportable during 1938 [13]. There are several distinct enzootic lineages of RABVs circulating in bats and wild carnivores in the USA [14], though most control and management efforts are focused on raccoon RABV [15]. The South Central Skunk (SCSK) variant of RABV was likely first detected in Texas as early as 1953 [16], although typing methods which could detect and identify this particular variant were not reported until 1986 [17]. Recent studies have documented that this variant of RABV has been expanding in geographic distribution [18] and recently invaded novel areas in the USA, causing epizootics in the northern part of Colorado for the first time [19]. This well-documented invasion presents an opportunity to quantify spatial emergence dynamics of skunk RABVs, and assess the validity of estimating spatial epidemiological parameters from surveillance data reported by the public.

Ecologists have developed occupancy models for quantifying species invasion processes [20], which are essentially the same phenomena as the emergence and geographic spread of novel pathogens. Occupancy is the process of a target species or pathogen being present or absent across space. Dynamic occupancy processes additionally consider a time component. Occupancy frameworks can incorporate an ecological process(es) of geographic spread and observation error separately, allowing for clearer interpretation of the influence of factors driving each process. Recently, occupancy models have begun to be applied to disease systems for estimating pathogen prevalence [21–23] but are under-utilized for quantifying parameters that describe the spatial spread of disease.

Using surveillance data from the SCSK epizootic in Colorado, we developed a dynamic patch-occupancy model to jointly estimate parameters describing the spatial spread of rabies while accounting for variable sampling effort. Considering surveillance data from dead skunks, we used the framework to: 1) quantify the rate of geographic spread and transmission distance per infection; 2) quantify effects of local infection density, seasonality and direction on the probability of geographic invasion; and 3) predict the occupancy probability and prevalence of RABV on the landscape through space and time. We also conducted a phylogeographic analysis of viral genetic data from a subset of the reported rabies positive skunks to validate the estimated rates of spatial spread and transmission distance that we estimated using the occupancy model. The occupancy framework we present is simple to implement and can be used for inferring rates and direction of geographic spread in a variety of emerging disease systems, providing basic insight for understanding disease emergence and practical insight for risk assessment and response planning.

Methods

Ethics statement

The decision and implementation of euthanasia of animals was conducted by county authorities, before samples were received for the current study (i.e., we had no role in this). For the current study, the State of Colorado Department of Natural Resources issued annual Scientific Collection Licenses (14SALV2060, 15SALV2060) in order for us to receive a subset of dead carcasses for testing.
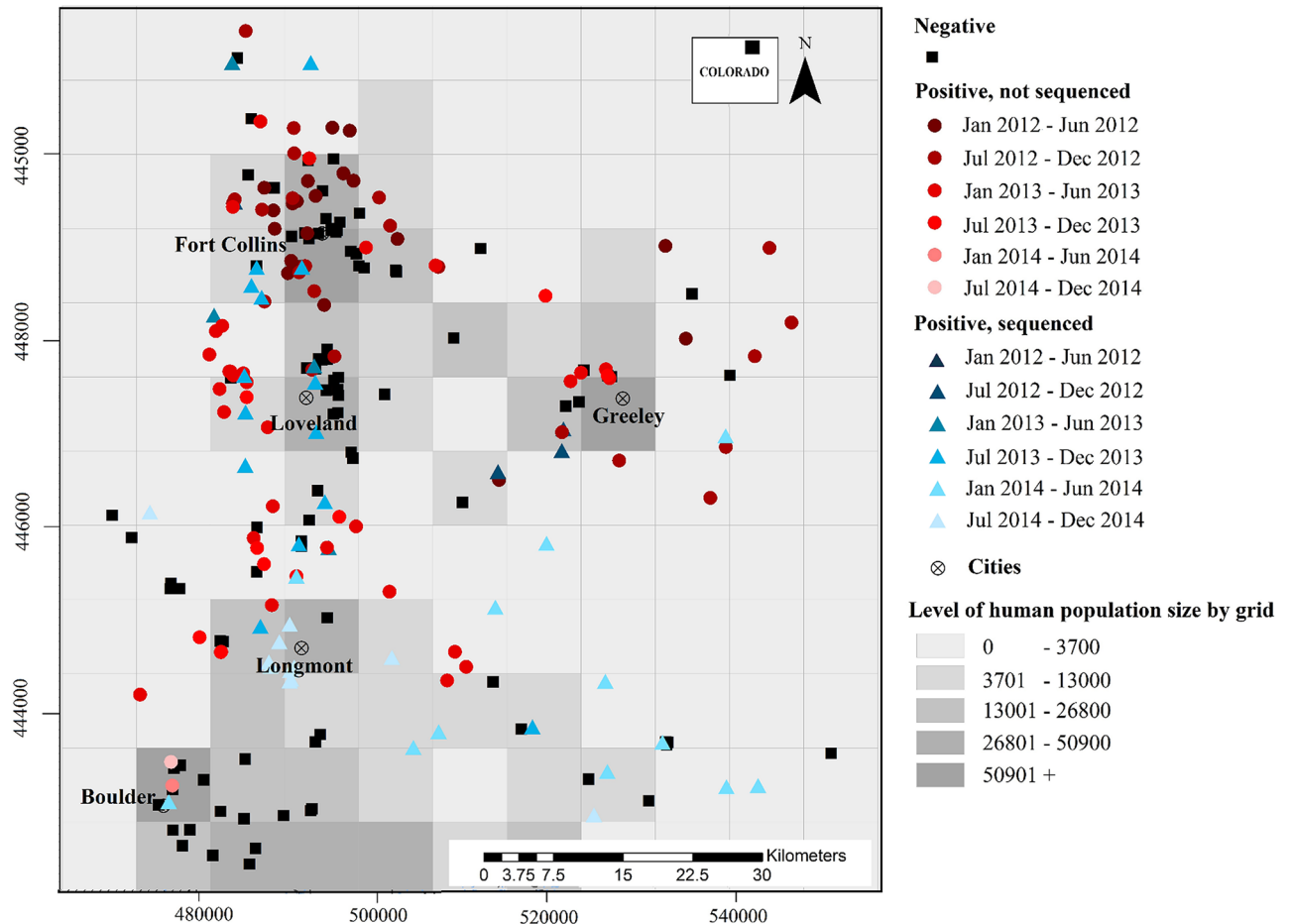
Fig 1. Map of study area. Each dot represents the location where a dead skunk was reported and subsequently tested for rabies. Grid cells show the scale at which the study area was gridded (8 x 8 km sites, i = 1, ..., M).

Surveillance data

The study area included 8448 km^2 (88 x 96 km; Fig 1) of Larimer, Boulder and Weld counties, Colorado, USA. Skunk samples were obtained by reports made by public to local health departments, which decided whether to submit a skunk for rabies testing to one of two diagnostic laboratories in the state. The decision and implementation of euthanasia of animals was conducted by county authorities, before samples were received for the current study. For the current study, the State of Colorado Department of Natural Resources issued annual Scientific Collection Licenses (14SALV2060, 15SALV2060) in order for us to receive a subset of dead carcasses for testing. There was no vaccination campaign (trap-vaccinate-release or oral rabies vaccination) targeting wildlife during the epizootic, but there were alerts through local public media and leash laws in effect.

Strange-acting or dead mammals found on the landscape with no reported human or domestic animal contact were not typically tested by the local health departments—especially later in the study period when the perceived risk of rabies infection was much lower and funding for testing had decreased. Because of the ongoing epizootic, these samples were referred for testing by United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services, National Wildlife Research Center (NWRC); hereafter referred to as

'enhanced surveillance'. Enhanced surveillance accounted for 0%, 1% and 47% of surveillance data in 2012, 2013 and 2014 respectively (Table SM1.1 in S1 Text). Enhanced surveillance samples were identified through the public health surveillance system by the local health departments, but with carcass referral to NWRC for rabies testing. Carcasses were referred to NWRC when the perceived risk to public health was low (i.e., due to a lack of contact with humans or pets). Because the rabies epizootic was waning in 2014, the high proportion of enhanced surveillance in 2014 relative to 2012 and 2013 did not contribute much extra information (except for improving uncertainty levels) using the occupancy model. Thus, typical surveillance systems could likely make the same type of inferences, but would show greater uncertainty as sampling decreased. An address matching-geocoding technique, using ArcGIS 10.3 (Environmental Systems Research Institute, Redlands, CA, USA), was used to convert street addresses of the case reports to UTM (Universal Transverse Mercator) coordinates.

Before 2012, carnivore rabies had not been detected in the study area for several decades. From 2012–2014 there were 246 skunk reports and 379 non-skunk terrestrial animal reports tested for rabies (Fig 1, Table SM1.2 in S1 Text), with a total of 139 rabies-positive skunks (raw prevalence ~ 57%, Table SM1.2 in S1 Text), and 21 rabies-positive cases from the 379 non-skunk terrestrial species tested (5.5% positive; Table SM1.2 in S1 Text) including: raccoons (6), bison (2), foxes (6) cats (2), horses (2), cows (2), coyotes (1). All non-skunk terrestrial animals were excluded from the analysis because they comprised such a small proportion of the samples and because surveillance efforts may have differed in some of these hosts—especially raccoons which are higher density and more peridomestic, and were experiencing an ongoing epizootic of canine distemper virus. Human population data from 2013 [24] were organized at the grid-cell scale (8 x 8km) to be included in the occupancy model as a potential factor contributing to reporting rate.

Occupancy model formulation

We used a discrete-time, dynamic, patch-occupancy model to quantify the occurrence probability of skunk rabies in space and time. In a typical occupancy framework, there is the partially observable occupancy process and an observation process that is conditional on the occupancy process. For this dynamic model the occupancy process describes the probability of rabies being present by site (grid cells) and time (months). The ecological process of rabies occurrence changed over grid cells and months due to the local component processes of colonization, local extinction, and persistence. Within the latent ecological process, we estimated important parameters describing the initial colonization dynamics (i.e., spatial invasion) of skunk rabies, specifically: direction of invasion, distance to nearest infected grid cell, and density of infection in the local neighborhood. Conditional on the latent ecological process of occupancy, we modeled the observation process as the probability that an animal sampled has rabies, given that rabies is present (prevalence given rabies occupancy). Within the observation process, we considered the effects of human population size on the proportion of skunks that were rabies positive. Thus, our approach allowed us to quantify parameters underlying the spatio-temporal dynamics of geographic spread while accounting for imperfect detection. For the occupancy model, we assumed the landscape was homogenous because our genetic model suggested that spread rates were homogenous across the landscape and we did not have enough degrees of freedom to incorporate more parameters in our occupancy model.

We used a monthly time step to scale with the incubation period for rabies virus in skunks [25,26]. We divided the study area (much of Larimer, Weld and Boulder Counties in Northern Colorado, USA) into 8 x 8 km grid cells ("patches" in occupancy modeling terms; Fig 1). This spatial scale was much larger than a skunk home range size [27,28] suggesting that local

transmission (i.e., due to skunk movements alone) should be primarily within grid cells or nearest-neighbor grid cells. Also, at this spatial scale, there were multiple samples collected within grids (Fig 1) at a given time step (mean number of samples in grids with samples: 3.6 sampled skunks, range 1–19 sampled skunks), which is necessary for estimating grid-cell level prevalence and informing the observation process, but at finer spatial scales multiple samples per grid cell and time step were rare. Thus, because of the sparse sampling intensity in our study area, the 8 x 8 km grid cell size was the smallest size we could choose for approximating the skunk home range (~ 2 x 2 km; [25, 26]) and hence skunk-to-skunk transmission distances.

Data on the presence of rabies in dead skunks (sampling unit = 1 skunk) were collected repeatedly in grid cells i = 1, . . ., M during each month t = 1, . . ., T (spanning January 2012 through December 2014). We modeled the true occupancy status $z_{i,t}$ conditioned on the previous time step $z_{i,t-1}$, as a latent Bernoulli variable defined by parameter $\Psi_{i,t}$ (occupancy probability), where $z_{i,t} = 0$ indicates that grid cell i is not occupied (no rabies in skunks) in time step t, and $z_{i,t} = 1$ indicates that grid cell i is occupied (at least one skunk with rabies) in time step t.

$$z_{i,t}|z_{i,t-1} \sim Bern(\Psi_{i,t}) \tag{1}$$

We assumed the initial occupancy state for each grid cell ($z_{i,1}$) was also a Bernoulli random variable described by occupancy probability $\Psi_{i,1}$, which had a Uniform prior distribution ($\Psi_{i,1} \sim Unif(0, 1)$). We allowed occupancy probability in grid cell i at time t ($\Psi_{i,t}$) to be determined by three types of local dynamic processes: initial colonization ($\gamma_{i,t}$), recolonization ($\zeta_{i,t}$) and persistence ($\phi_{i,t}$), which were conditional on the latent occupancy status in the previous time step ($z_{i,t-1}$).

$$\Psi_{i,t} = \gamma_{i,t-1}(1 - z_{i,t-1})(1 - A_{i,t-1}) + \zeta_{i,t-1}(1 - z_{i,t-1})A_{i,t-1} + \phi_{i,t-1}z_{i,t-1} \tag{2}$$

We used parameter $A_{i,t}$ to distinguish an initial colonization event from subsequent colonization events. Although it is more typical to consider only occupancy and extinction dynamics in occupancy models, we further distinguished initial colonization from recolonization (as in [29]) because we were most interested in factors (described below) driving spatial invasion into new sites. We quantified the effects of multiple different factors (direction, distance to nearest infected neighbor, neighborhood infection density (Eqs 3–8) on initial colonization probability ($\gamma_{i,t-1}$), and treated the other processes as inherent contributors to occupancy probability ($\Psi_{i,t}$) without quantifying their potential drivers (i.e. we used a global parameter for persistence- $\phi_{it} \sim Beta(\alpha_\phi, \beta_\phi)$ and recolonization—$\zeta_{it} \sim Beta(\alpha_\zeta, \beta_\zeta)$. For $\gamma_{i,t-1}$, an intercept term (β_o) absorbed non-spatial effects (Eq 3). Direction (north-south (N) or east-west (E)) was modeled as increasing integers (1 representing the southernmost grids and 12 representing the northernmost grids for N; and 1 representing the westernmost grids and 11 representing the easternmost grids for E; Eq 4, β_k, where k = 1, . . ., K for the number of initial colonization effects considered). East-west direction was represented similarly but using columns in the grid. We modeled an interaction between direction and a trend in time (T) by multiplying the direction covariate data by the time step ($N_i \cdot T$ or $E_i \cdot T$). We calculated distance to the nearest infected grid cell by taking the minimum distance for all pairwise distances between the centroid of target grid cell i^* in time t and all other grid cell centroids (i) in time $t-1$ (d_{ii^*}). We estimated the relationship of initial colonization probability and distance using an exponential decay function which included a parameter describing the decay rate of initial colonization probability (α) with distance and a scaling parameter relative to maximum initial colonization probability (β_k, Eq 5). We estimated the local neighborhood infection density for grid cell i^* in time t using the occupancy prevalence in all immediate grid-cell neighbors (queen's neighbors;

$j = 1, \ldots, J$; note j's are a special subset of i—restricted to the closest neighbors of i^*) in time step $t\text{-}1$ and estimated its effect with parameter β_k (Eq 6). We incorporated a seasonal effect as a factor where one level represented spring/summer (Feb.–Aug.) and the second level represented fall/winter (Sept.–Jan.) (Eq 7). We chose these time frames because the literature and surveillance data suggest peaks of rabies in spring and summer following arousal and mating activities, and subsequent peaks in fall and winter associated with dispersal and contact involving susceptible young of the year [16, 25]. We modelled the infection density (Eq 6) and distance effects (Eq 5) separately because we were interested in the separate effects of distance versus infection intensity. Together these effects describe the spatial kernel for local transmission. Additionally, we modelled a weighted spatial kernel (Eq 8) which accounted for the density of neighbors with infections but discounted the impact an infected cells at further distances from the grid of interest (d_{ii^*} represents the distance between grid cell 'i' and the grid cell of interest 'i^*').

$logit(\gamma_{i,t\text{-}1}) =$

$$\beta_0 \quad \text{(non spatial)} \tag{3}$$

$$\beta_1 N + \beta_2 T + \beta_3 N * T \quad \text{(direction by time)} \tag{4}$$

$$\beta_4 e^{-\alpha min(d_{ii^*})} \quad \text{(transmission distance)} \tag{5}$$

$$\beta_5 (1/J) \sum\nolimits_{j=1}^{J} [z_{j,t-1}] \quad \text{(local neighborhood effect)} \tag{6}$$

$$\beta_6 S \quad \text{(season} - \text{Feb.} - \text{Aug. vs Sept.} - \text{Jan.)} \tag{7}$$

$$\beta_7 \sum\nolimits_{zi,t-1=1} [1/d_{ii^*}] \quad \text{(weighted spatial kernel)} \tag{8}$$

We considered these effects (Eqs 4–8) separately and additively (up to three effects including the Eq 3) in our model selection procedure (described below). We did not present the fullest model because parameter estimates became inconsistent when more than three additional effects were in the model. Persistence and recolonization were important processes in the occupancy dynamics, thus we modeled them explicitly using Beta prior distributions with shape and scale parameters ($\alpha_\phi = 1$, $\beta_\phi = 1$, $\alpha_\zeta = 1$, $\beta_\zeta = 1$). We did not include covariates that could potentially drive persistence and recolonization because we were only interested in the process of spatial spread (i.e., initial colonization). Prior distributions for all β_k parameters were: $\beta_k \sim$ Norm(0, 1), except for the scaling parameter for the exponential decay model (Eq 5) which was modeled as a Gamma(5, 1).

For the observation layer of our model, we represented the number of skunks that were positive for rabies, y, in grid cell i at time t when rabies was present as the observed data using a binomial distribution where p was the estimated prevalence of rabies in dead skunks (given occupancy) and the number of trials ($R_{i,t}$) were the observed number of samples collected in grid cell i at time t.

$$y_{i,t} = \{0, for\ z_{i,t} = 0, i = 1, \ldots, M; Bin(p, R_{i,t}), for\ z_{i,t} = 1, t = 1, \ldots, T\} \tag{9}$$

To account for variation in rabies detection due to variation in human population size, we examined prevalence as a linear function of human population size (N).

$$logit(p_i) = \eta_0 + \eta_1 log_{10}(N_i) \tag{10}$$

We only investigated this effect in the intercept-only model (Eq 3) and the two-effect model we were most interested in (Table SM2.1 in S1 Text for list of all models that were fit). The general model specification is given in section SM2.1 in S1 Text.

To calculate the posterior distribution for the parameters of interest, we fit models using a Markov chain Monte Carlo (MCMC) algorithm with a Gibbs sampler including Metropolis-Hastings steps [30] custom written in program R [31]. Posterior estimates for the data are based on 50,000 iterations of the MCMC algorithm with the first 5,000 iterations discarded as burn-in. Convergence and mixing were assessed graphically. Example MCMC chains and posterior distributions are given in SM2.1 in S1 Text. Convergence of the best predictive model was additionally assessed using the Gelman-Rubin statistic [30]. The joint and conditional distributions are specified in SM2.3 in S1 Text. We evaluated the ability of our models to recover parameters by simulating data from known parameter values and estimating the parameter values using our fitting framework (Model Validation Method in SM5 in S1 Text).

Model selection and evaluation

We used a combination of Watanabe Akaike Information Criterion (WAIC) [32, 33], Area Under the receiver operator Curve (AUC) [34], and leave-one-out cross validation (looCV) [31] to compare the importance of different effects on γ and p and assess goodness-of-fit (presented in Table 1, schematic of work flow shown in Fig. SM3.1). WAIC is a model selection criterion based on the posterior predictive distribution, and was used to compare fits of models. WAIC was not used to compare models with human population modelled on prevalence (p) because the data were different and thus the WAIC values would not be comparable. AUC is a measure of how well variation is explained by the model—in our case, the ability to distinguish a presence from an absence—and was used to assess how well a particular model explained the data (i.e. a measure of goodness of fit as in [35]). LooCV is a measure of the model's ability to predict out of sample, and thus was used to compare predictive ability among models. When comparing predictive ability between models we used both AUC and looCV.

We calculated WAIC and AUC (Section SM4 in S1 Text) for fits to the full data as well as the looCV predictions. WAIC is preferable to DIC (Deviance Information Criterion—another Bayesian method for model selection) for model selection using hierarchical models [32] because it considers the posterior predictive distribution explicitly and penalizes for complexity of model structure, not just the number of parameters. Similar to DIC, lower values of WAIC indicate a better model of the data. However, in contrast to DIC [36], there is no standard quantitative difference between WAIC values from alternative models that indicates a significant difference between them (i.e., the only criterion is that lower is better). We presented two AUC scores: 1) AUC1 measured predictive ability of p using predicted y's from posterior values of z and observed y's, 2) AUC2 measured predictive ability of Ψ using the posterior values of z and the observed y's transformed to binary data. For the out-of-sample predictions, we conducted looCV for each point in the data (using all other points as the training data) and presented means of AUC1 and AUC2 as measures of predictive ability.

Out-of-sample prediction

To further evaluate the predictive ability of our approach we predicted occupancy status over space and time using only parameters estimated from the "best predictive model" (i.e., considering out-of-sample statistics for AUC1, AUC2 and looCV) and the sample size data. First, we fit the best model to the full data and then predicted all $y_{i, t}$ from the posterior predictive distribution [30]. Each predicted $y_{i, t}$ depended on the predicted $y_{i, t-1}$, rather than the data, but the

Table 1. Model selection and fit statistics for within- and out-of-sample predictions.

#	Effects on initial colonization probability (γ)	p	Within sample			Out of sample		
			WAIC	AUC1	AUC2	AUC1	AUC2	looCV
1	·	-	0.804	0.95	0.64	0.78	0.62	-205.71
1b	·	Pop	NA	1	0.64	0.81	0.62	-218.22
2	Distance	-	0.803	0.96	0.7	0.78	0.69	-205.72
3	NxT	-	0.803	0.96	0.66	0.78	0.66	-205.72
4	ExT	-	0.804	0.95	0.66	0.78	0.67	-205.72
5	Neighborhood	-	0.803	0.96	0.72	0.78	0.7	-205.71
6	Season	-	0.802	0.95	0.71	0.78	0.7	-205.72
7	Kernel	-	0.804	0.96	0.72	0.78	0.7	-205.72
8	Distance + NxT	-	0.804	0.96	0.68	0.78	0.7	-205.95
9	Distance + Neighborhood	-	0.804	0.96	0.76	0.78	0.74	-205.99
10	NxT + Neighborhood	-	0.804	0.96	0.71	0.78	0.69	-205.79
11	Distance + Season	-	0.804	0.96	0.76	0.78	0.75	-205.77
12	NxT + Season	-	0.804	0.96	0.66	0.78	0.71	-205.84
13	Neighborhood + Season	-	0.804	0.95	0.73	0.78	0.71	-205.72
14	Kernel + N*T	-	0.804	0.96	0.71	0.78	0.69	-205.79
15	Kernel + Season	-	0.804	0.97	0.75	0.78	0.68	-205.99
16*	Neighborhood + Season + Distance	-	0.804	0.96	0.76	0.78	0.73	-205.88
9b	Distance + Neighborhood	Pop	NA	0.99	0.75	0.8	0.74	-218.7
11b**	Distance + Season	Pop	NA	0.99	0.78	0.81	0.72	-219.71
16b	Neighborhood + Season + Distance	Pop	NA	0.99	0.74	0.8	0.72	-218.05

*Model used to estimate effects in Fig 2;

**Best predictive model, used to make predictions in Figs 4 and 5; · = intercept only, - = no effect, Pop = log human population size effect on prevalence parameter p, Distance = effect of distance to nearest infection, Neighborhood = local neighborhood infection density effect, N·T = north-south by time directional effect, Season = factor with two levels (spring/summer: Feb.-Aug. and Fall/Winter: Sept.-Jan.), E·T = east-west by time directional effect, -WAIC = Watanabe Akaike Information Criteria (section SM4.1 in S1 Text), AUC1 = Area Under the Curve for observed positive counts (section SM4.2 in S1 Text), AUC2 = Area Under the Curve for latent occupancy process (section SM4.3 in S1 Text), looCV = leave-one-out cross validation score (section SM4.4 in S1 Text), NA = not available; we did not display these WAIC values because they cannot be compared to the WAIC values from the models without Pop.

actual data for sample size ($R_{i,\,t}$) were used in the prediction in order to scale $y_{i,\,t}$ predictions appropriately.

Rates of spatial spread from occupancy model

Using the model with the best out-of-sample looCV score, we predicted grid cell occupancy over time ($z_{i,\,t}$; rabies presence -1, or absence -0), and estimated the rate of southerly spread of the predictions using regression. First, we obtained the $z_{i,\,t}$ values from each MCMC iteration (minus the burn in) for grids across time. Each grid cell corresponded to a distance in kilometers from the southernmost border of the study area (e.g., 1, . . ., 88 km; "North value"). For $z_{i,\,t} = 1$ values, we ran a simple linear regression where the independent values were the months in time and the dependent values were the north values. The slope of this regression represents the monthly rate of southerly spread. We then converted this to an annual rate of southerly spread by multiplying the monthly rate by 12. We calculated the annual variance using the delta method [37].

Genetic sequencing

To verify spatial patterns indicated by our occupancy model, we sequenced the whole or partial glycoprotein (G, 1575 base pairs [bp]) and the non-coding region between the glycoprotein and polymerase genes (GL, 560 bp) for 53 viruses collected in Colorado between August 2012 and December 2014 (Table SM1.3 in S1 Text). We added sequence data from 20 viruses collected through May 2015 (27.4% of the total genetic dataset) to increase the precision of parameters of the molecular evolutionary models. We confirmed that these additional sequences comprised an extension of the same epizootic though preliminary phylogenetic analysis and the absence of changes in the inferred rate of spread in 2015 which could have biased our overall spread rate. Rabies virus RNA was extracted from brainstem tissue samples of rabies positive skunks using Trizol reagent following the manufacturer's protocol. The conversion of RNA to cDNA (RT) and primary PCR amplification was accomplished using a Superscript III One-step RT-PCR system with Platinum Taq DNA Polymerase (Invitrogen), targeting a 2,135bp region including the full glycoprotein gene with previously published primers [18,38]. PCR products were visualized by UV-light on a 2% agarose gel with ethidium bromide, and cleaned using ExoSap-IT (Affymetrix) following the manufacturer's protocol. Sequencing reactions were performed using Big Dye Terminator v.3.1 (Applied Biosystems), with flanking and internal primers as previously described [18,38]. Sequencing products were cleaned using Sephadex G-50 columns (GE Healthcare) and run on a 3130 or 3500 analyzer (Applied Biosystems). Forward and reverse sequences were aligned in Sequencher v.5.2.4 (Gene Codes Corporation), and ambiguities were resolved visually. Alignment of full or partial glycoprotein sequences was performed using BioEdit v.7.2.0.

Phylogeographic analysis of viral spread

Continuous phylogeographic analysis was conducted in BEAST v. 1.8.4 [39]. Briefly, this method estimates the phylogenetic history connecting samples and conducts ancestral state reconstruction of the latitudes and longitudes of inferred nodes using time and location-annotated sequence data. Phylogenetic uncertainty is accommodated by summarizing estimates across the posterior distribution of trees from a Bayesian search. Analyses in TempEst (software that analyzes correlations between the temporal and genetic distances between sequences) showed evidence of a molecular clock signal in the coding (G) and non-coding (GL) regions of the RABV genome, but faster evolution in the G-L region compared to G [40]. Our BEAST analysis therefore modeled a single tree topology (given that the samples represent the same underlying epidemic history) using different molecular clock and substitution models to allow for differences in evolutionary rates in different parts of the RABV genome. Preliminary BEAST runs using the lognormal relaxed molecular clock indicated little variation in rates among branches, indicating the use of the strict molecular clock for both partitions, which was supported by equivocal differences in Bayes Factors (BF) between models assuming strict or relaxed molecular clocks (BF = 1.5 in favor of strict clocks) [41]. Final runs used customized substitution models for the following data partitions: G codon positions 1+2 = TPM1uf; G codon position 3 = TIM1+G, GL = TPM1uf+I, as suggested by AIC in jModeltest2 [42]. Among the 3 random walk phylogeographic models tested (homogenous, Cauchy and gamma), marginal likelihoods estimated by the stepping stone method were highest in the homogenous model (-3574.86) followed by the gamma (-3577.27) and Cauchy (-3594.05) models, suggesting relatively little variation in spread rates among branches or little power to detect such variation. To account for the possibility of among branch variation, we present the results of the gamma model, but note that parameter estimates were nearly identical in the similarly supported homogeneous rate model. All analyses used the Bayesian skyline model of

demographic growth. MCMC chains were run for 100 million generations which generated effective samples sizes >200 after removal of burn-in.

We used the Seraphim package of R to calculate viral diffusion rates from 500 randomly selected (post burn-in) trees from the posterior distribution of the BEAST analysis [43]. We estimated viral spread rates as (i) the "mean branch velocity" calculated as the average velocity across the branches of each tree, averaged across all trees and (ii) the weighted "diffusion rate", obtained by summarizing distances and times spanning each tree and taking the average of that value across all trees (i.e., sum distances/sum time lengths over entire tree). To calculate the transmission distance per viral generation (i.e., per infected animal), we divided the distance traversed along each branch by the expected number of infections along that branch, assuming a generation time of 30 days, which corresponds to the incubation period of rabies virus in skunks [25,26]. However, the substitution rates that we estimated (G: 4.1×10^{-4} substitutions per site per year (95% highest posterior density (HPD) = $2.3–6.05 \times 10^{-4}$); GL: 8.7×10^{-4} [HPD_{95} = $4.1–13.4 \times 10^{-4}$]), imply that some transmission events would be expected to occur without detectable evolution in the partial genomes that we sequenced. This led to branch lengths that were less than the assumed generation time of RABV, which caused an upward bias in inferred transmission distances per infection. We corrected for this bias by forcing all branches to have a minimum generation time of 1 infection.

Results

Model selection

The probability of initial colonization (γ) was best explained by the minimum distance to infections, infection density in the local neighborhood, season, and spatial kernel (see AUC2 of Models 1–7; Table 1). Similarly, the best two-effect models included pairs of these effects (distance + neighborhood, distance + season, kernel + season; Models 9, 11, 15; Table 1). Accounting for direction (i.e., N·T or E·T) provided more minor improvements to quantifying initial colonization probability (AUC2 Table 1). The inclusion of human population size on p (rabies prevalence in occupied grid cells) significantly improved predictive power of the model (see AUC1 and looCV in Table 1, Models 1b, 9b, 11b, 16b) because it explained some of the prevalence variation due to differences in the number of samples reported by humans. We considered Model 11b to be the best predictive model because it had the lowest looCV, and we used this model for prediction. However, in order to study the effects of the most significant single predictors (neighborhood, season and distance), we fit Model 16 to quantify the relative importance of these factors (results presented in Fig 2).

Effects on initial colonization probability

We modeled the effect of distance to nearest infection using an exponential decay function. Using the estimated decay rate parameter, we calculated the distance at which initial colonization probability decayed to half its maximum (referred to as "transmission distance") using the asymptotic limit as the minimum (see section SM54 in S1 Text for calculation). We interpreted the transmission distance as a measure of skunk-to-skunk transmission distance. Transmission distance was 3.9 km (95% credible intervals (CI_{95}): 1.4, 11.3; Fig 2A). Similarly, using the genetic data independently, the mean distance per viral generation (another proxy for skunk-to-skunk transmission distance) was estimated to be 2.3 km (95% Highest Posterior Density (HPD_{95}) = 0.04–5.7; Fig 3C).

Initial colonization probability increased exponentially with local neighborhood infection density, especially after 0.2 (i.e., > 1grid cell occupied; Fig 2B, Model 16, Table 1). Initial colonization probability was substantially higher during the spring/summer season relative to the
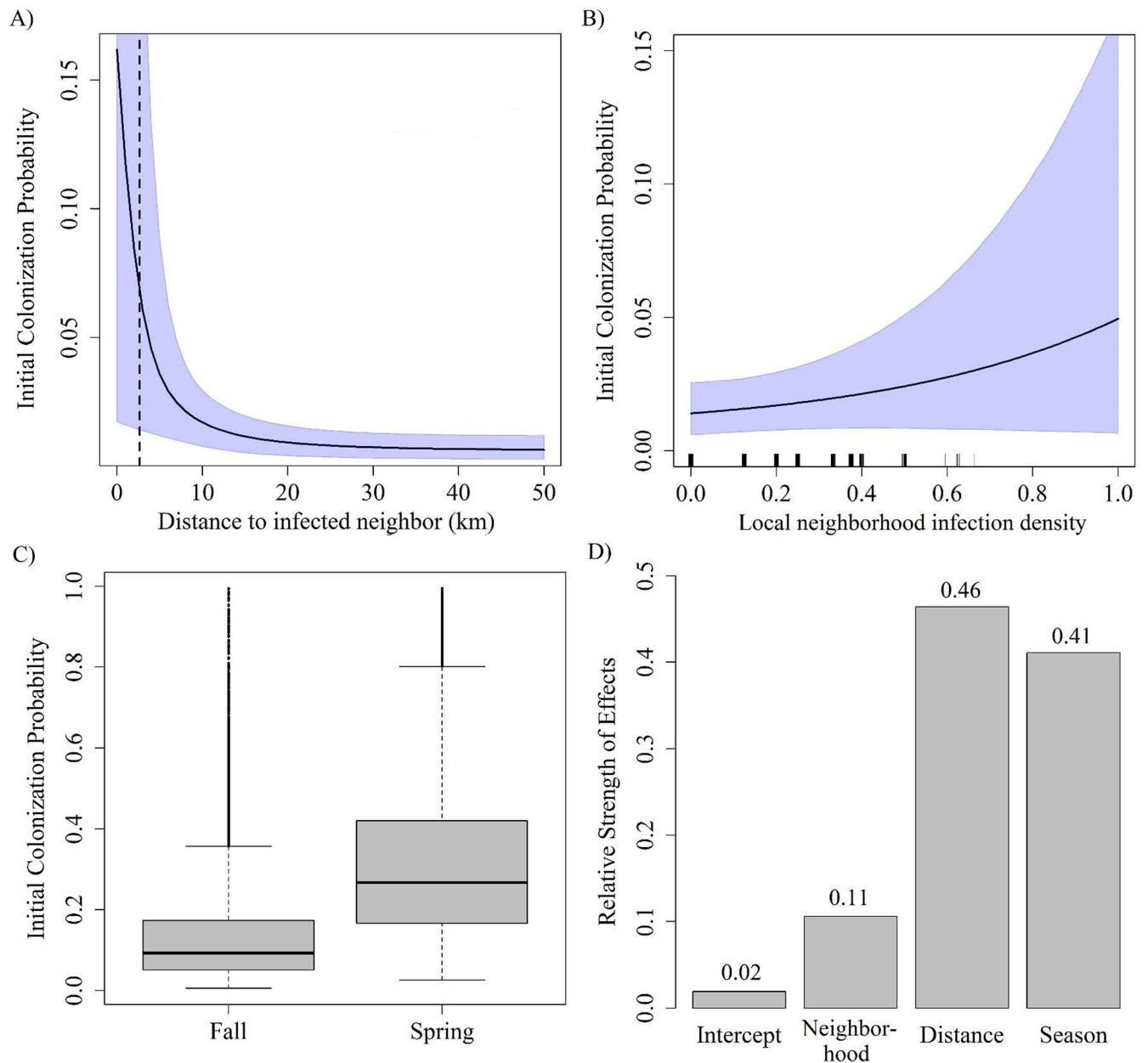
Fig 2. Effects of covariates on initial colonization probability. Predicted γ as a function of different factors using Model 16 Transmission distance = 3.9 km 95% CI (1.4, 11.3) (neighborhood+distance+season, Table 1). A) Decay of initial colonization probability with distance to nearest infected grid cell in km. B) Proportion of local neighborhood ("queen's neighbors") infected. C) Season factor with two levels. Median values of initial colonization probability are indicated by the horizontal black line. D) Relative contribution of different effects (see section SM5 in S1 Text for calculations). Shading in A and B indicates 95% credible intervals.

fall/winter season (Fig 2C, Model 16). Of the 76% of variation in occupancy probability that was explained by the three-factor model (AUC2, Model 16, Table 1), 57% of initial colonization probability was explained by local effects (distance or neighborhood), 41% was explained by season and 2% was explained by other factors (which could include translocation or other unknown factors; Fig 2D). Although direction explained less variation in initial occupancy probability relative to other factors (Table 1, Models 3 and 4 versus Models 2, 5, 6, and 7),
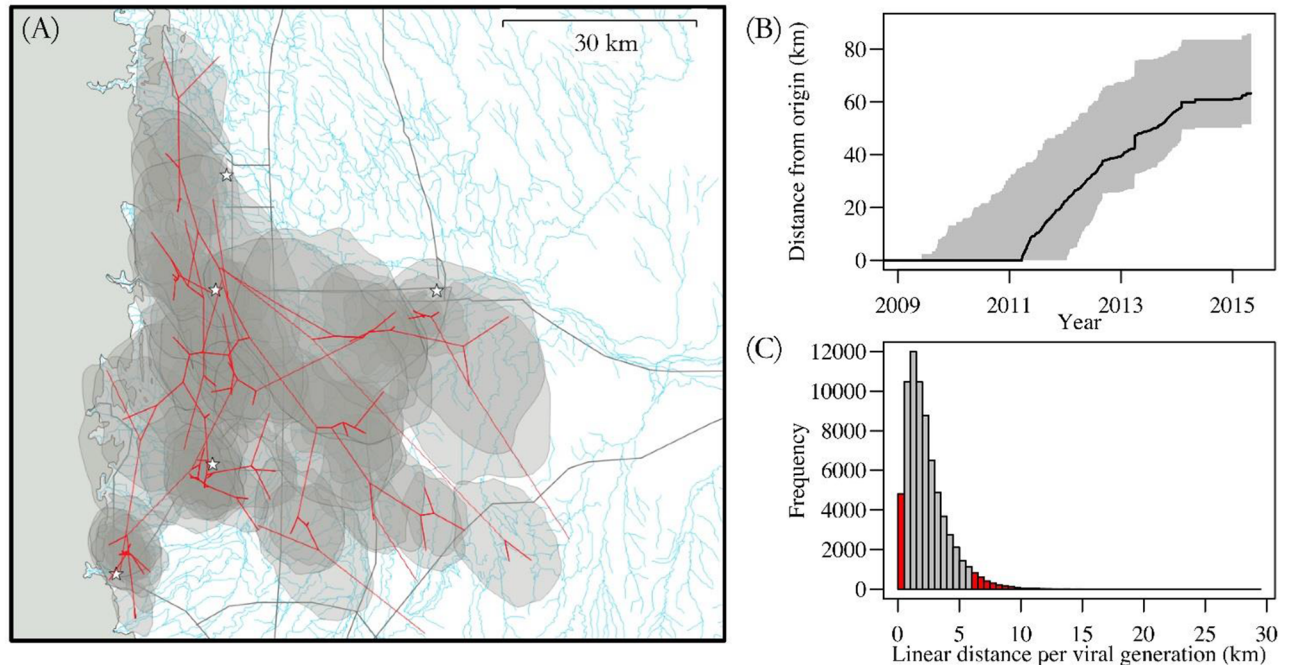
Fig 3. Phylogeographic inference of viral spatial dynamics. A) Spatio-temporal projection of the maximum clade credibility tree from the phylogeographic analysis. Transparent polygons indicate the 80% HPD of infected area through time. Grey shading on the left indicates landscape that is above 1829 m. Grey lines indicate major highways. Blue lines indicate all waterways, including very minor and ephemeral ones. Stars indicate the centers of major cities (as in Fig 1). B) Spatial expansion of RABV in skunks, displayed as the distance from the inferred epizootic origin in the phylogeographic analysis. Grey shading indicates the HPD_{95}. The solid black line is the median distance. C) The distance traversed by each branch in the phylogenetic tree divided by the expected number of infections along that branch, assuming a generation time of 30 days. The histogram shows distances calculated from the 144 branches of 500 randomly-selected trees from the posterior distribution of the phylogeographic analysis. Grey bars are the inner HPD_{95}.

https://doi.org/10.1371/journal.pntd.0005822.g003

there appeared to be a significant increase in initial colonization probability in the southerly direction with time during the study period indicating southerly spread of the disease (Fig SR1.1 in S1 Figures). However, east-west movement over time did not show significant visual trends (Fig SR1.2 in S1 Figures).

Prediction

The "best" predictive model we examined (Model 11b, Table 1) performed very well at predicting rabies cases in space and time using parameters trained on different data than the data being predicted (Table 1, out-of-sample prediction AUC1 & AUC2). The model also performed very well at out-of-sample prediction using only the initial conditions, sample size data and parameters fit to the full data (Fig SR2.1 in S1 Figures). Average occupancy probability across the study area was low over time (Fig 4, top row, mean: 0.034, CI_{95}: 0.006–0.24) but prevalence in sampled dead skunks in occupied grid cells was high (Table 2, mean: 0.91, CI_{95}: 0.86, 0.95). Occupancy probability was predicted to decrease substantially in the north and increase in the south during the time course of the study period (Fig 4, Figs SR3.1 & SR3.2 in S1 Figures). Also, earlier during the time course occupied patches were larger and more contiguous whereas later occupied patches became more fragmented and isolated (Fig 4). The overall rate of southerly spread estimated by the occupancy model was predicted to be 21.1 km/year (CI_{95}: 16.7, 25.5, Fig 5). This was similar to the phylogeographic diffusion rate assuming either gamma 21.8 [HPD_{95}: 14.9–29.0] or homogenous rate 22.6 [HPD_{95}: 15.3–29.7]
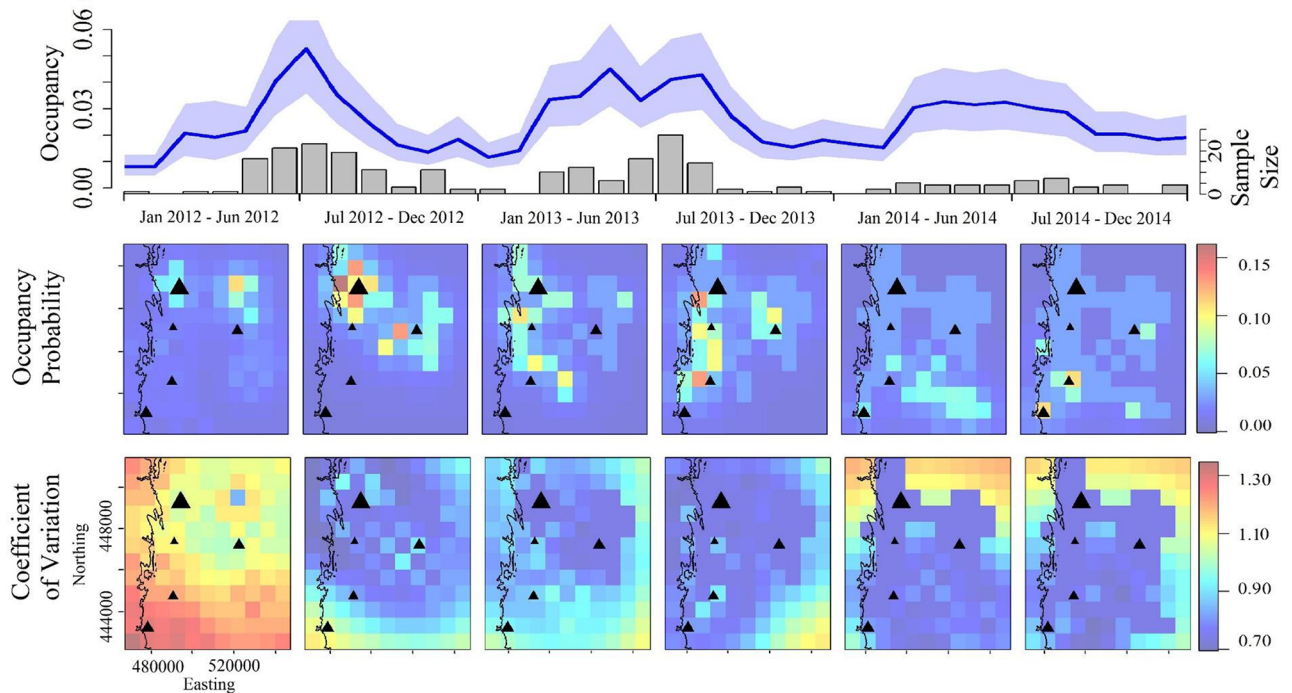
Fig 4. Rabies occupancy probability over time. Top. Occupancy probability for all sites over time (blue line, shading: 95% credible intervals) using Model 11b, Table 1). Grey bars: number of samples collected. Middle. Occupancy probability in space and time. Bottom. Coefficient of variation for occupancy probabilities over space and time. Middle and bottom: Data were aggregated over a 6-month time frame (corresponding to the X-axis labels of the top row plot). Black triangles represent city locations (Fort Collins, Greeley, Boulder and Longmont); sizes scaled to the human population size of the cities. Topographical divide from Fig 1, which indicates land above 1829 m (6000 feet), is indicated by the black line.

https://doi.org/10.1371/journal.pntd.0005822.g004

models. Mean branch velocities from the phylogeographic models were slightly higher but had overlapping HPD$_{95}$s with the genetic and occupancy-based models (gamma: 28.4 [HPD$_{95}$: 19.6–39.8] or homogenous 29.8 [HPD$_{95}$: 20.8–39.8]).

Discussion

We used an occupancy modeling framework to understand and predict the spatial spread dynamics of an important zoonotic disease in a reservoir host species. The framework produced good out-of-sample predictions of spatial dynamics despite our relatively small dataset, suggesting it can be useful for predicting spread in other surveillance systems that rely on passive surveillance data. We converted occupancy probabilities into rates of directional spread, which is useful for planning interventions and allocating resources for disease management. The rates of spatial spread predicted by the occupancy model appeared similar to those predicted by phylogeographic analyses. Thus, our phenomenological approach for estimating rates of spatial spread may provide accurate prediction of spatial spread rates based on parameters estimated in the recent past for a particular region. While the phylogeographic method incorporates additional information regarding connectivity between cases (i.e., genetic sequence similarity), the occupancy model produces similar results with only the spatial and temporal information. The phylogeographic method has the advantage of not requiring negative samples, while the occupancy approach has the advantage of not requiring genetic data. Each may therefore be suited to address alternative data gaps typical to surveillance systems.

Table 2. Parameter estimates from the best predictive model of occupancy probability (Model 11b, Table 1) and the phylogeographic model.

|	Parameter Estimates (uncertainty)
Occupancy probability	0.034 (0.006, 0.236)
Persistence probability	0.22 (0.15, 0.31)
Initial colonization probability	0.018 (0.01, 0.06)
Re-colonization probability	0.032 (0.02, 0.04)
Prevalence	0.91 (0.86, 0.95)
AUC1	0.96
AUC2	0.76
Local neighborhood parameter (β_κ)	1.15 (-0.53, 2.86)
Increase in γ for 1 additional infected neighbor[a]	0.002 (-0.000, 0.016)
Mean transmission distance in km[a] **(occupancy)**	3.9 (1.4, 11.3)
Mean transmission distance in km (phylogeographic)	2.3 (0.02, 5.7)
Rate of southerly spread in km/year (occupancy)	21.1 (16.7, 25.5)
Mean branch velocity (phylogeographic)	28.4 (19.6–39.8) (gamma) 29.8 (20.8–39.8) (homog)
Diffusion rate (phylogeographic)	21.8 (14.9–29.0) (gamma) 22.6 (15.3–29.7) (homog)
Total number of positives	140

Rate of southerly spread = slope of relationship between kilometers moved south and time Fig 5); Numbers in brackets (uncertainty) indicate 95% credible intervals for occupancy results and 95% highest posterior density for phylogegraphic results;

[a]See section SM5 in S1 Text for derivation of these values; Gamma and homogenous (homog) refer to different assumptions about the level of branch heterogeneity in the phylogeographic models. Models using Cauchy distributed rates were less supported (BF = 16.8 and 19.2 in favor of homogeneous and gamma, respectively) and are omitted from the table.

https://doi.org/10.1371/journal.pntd.0005822.t002

Both the phylogeographic and occupancy approaches can be sensitive to sampling intensity and scale, such that interpretation of results should consider sampling design.

Our approaches also present an alternative to detailed mechanistic models of disease transmission which may require more data than are available, such as knowledge of host population sizes and connectivity (e.g., [44]) (although we did not have host demographic data to explicitly test for differences between our phenomenological approach and a more mechanistic approach). The occupancy framework is flexible for testing the role of different contact heterogeneities. Any desired connectivity pattern could be incorporated by modifying the spatial terms using another description of spatio-temporal connectivity between rabies-positive cells. Similarly, if geographic barriers are important [45], or if there are data on ecological processes of known importance such as host density, additional terms could be added to occupancy processes to quantify these effects and improve prediction. In addition, some surveillance systems (e.g., the National Rabies Management Program of the United States Department of Agriculture, Animal and Plant Health Inspection Service, Wildlife Services) are comprised of multiple types of surveillance data (e.g., passive methods such as: public health reports, road kill reports, landowner reports of nuisance animals, or active methods such as: trapping), which each introduce their own sampling biases. The occupancy framework is flexible for incorporating multiple methods of surveillance simultaneously and, in the observation model, allowing for quantification of effects due to the type of surveillance data, which can in turn inform the
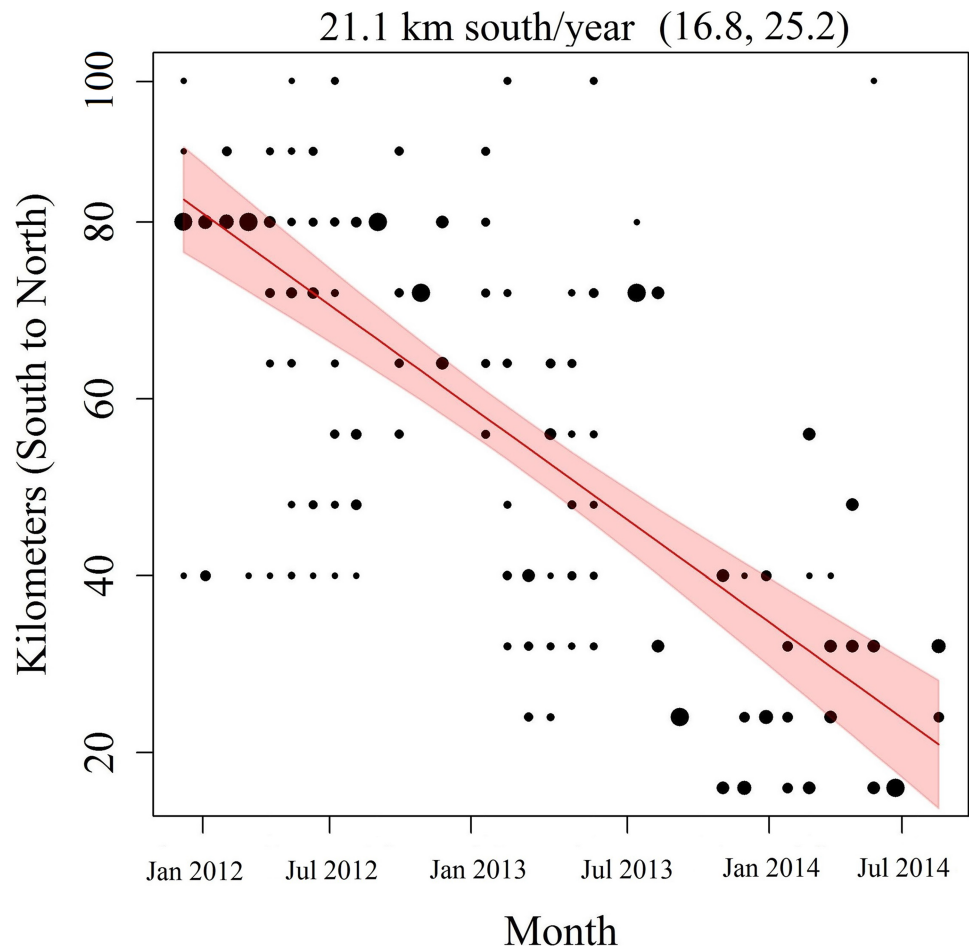
Fig 5. Rates of spatial spread. Model 11b was used to predict spatial spread of rabies. Points are the predicted distance in km of new colonizations from the southern-most grid-cell row. Size of the points corresponds to numbers of new colonizations at a given distance from the southern-most row. The slope (red line) gives a monthly rate of southerly spread (21.1 km/yr). Shaded area gives the 95% confidence intervals of the slope. The 6-month intervals shown on the X-axis correspond to those in Fig 4.

design of surveillance. In situations where vaccination or other interventions are applied, effects of the intervention could be quantified by incorporating those data as a covariate in the latent ecological process. Thus, our framework could be used to identify the magnitude of intervention effects and facilitate planning of where additional interventions are needed.

The directional rates we estimated by both the occupancy and phylogeographic models were twice as high as those estimated for SCSK RABV by [18] using a phylogeographic approach (10 km/year; range 4–12 km) on data from south-central USA. One reason for this discrepancy could be that the analysis of [18] averaged rates over a long time period and large geographic area (multiple US states), whereas our study covered a short time period within three counties in a single state. From a level IV ecoregion perspective, our study site included only 4 distinct ecoregions (Rolling sand plains, Flat to rolling plains, Front range fans and Foothill shrublands), while that of [18] included 50–100 ecoregion types. Thus, the landscape within our study area was less heterogeneous than the landscape within the study of [18], which could partly explain differences in rates of spatial spread of rabies. Also, estimates from longer time scales may have included re-emergence dynamics in addition to initial invasion

dynamics (reducing inferred rates of spatial spread). However, the rates we estimated are significantly lower than those found during the initial emergence of RABV in raccoons in the eastern USA (~38 km/year) [43,46], suggesting that skunk ecology, at least in south and western USA, leads to slower geographic spread of RABV. This is consistent with a previous study which used landscape resistance models and showed that skunks and raccoons used different movement corridors on the landscape [47]. Also, skunk rabies variants are 8.1 times less likely to be transmitted to foxes relative to raccoon rabies variants [14]–their lower propensity to spillover into longer-ranging hosts could contribute to slower spatial spreading.

Our mean estimates of individual-level transmission distance (~ 3.9 km by the occupancy model and 2.3 km by the phylogeographic model) were higher than the mean transmission distance estimated for dogs in the Serengeti, Tanzania (0.88 km) [48]. This estimate, which was based on infection reports, suggest a much closer mean transmission distance than was estimated in study of canine rabies in South Africa which was conducted using genetic data [mean distance between the most probable linked cases = 14.9 km; [49]. The latter study suggests a significant influence of anthropogenic movement of dogs, which is unlikely for skunks, and no rabies-positive dogs were reported. Although there were some canine species (6 foxes and 1 coyote) found to be positive for SCSK in our study area, these reports only comprised 5% of the rabies-positive reports suggesting that these longer-ranging hosts may only have made minor contributions to spatial spreading. Thus, our estimates of mean transmission distance likely reflect distances at which skunks move naturally during the contact and transmission process. Our approach required that multiple samples were collected in the same grid cell at the same time step in at least some of the grid cell/time steps (here we had a total of 180 grid cells with > 1 sample in the same time step over a total of 132 grid cells x 36 time steps = 4752 (3.8%)). Based on this requirement, we used a larger spatial scale (8 x 8 km) than skunk home range size (~ 2 x 2 km, would have been preferable). Thus, our estimates of transmission distance were likely biased high. For example, the transmission distance estimated in the occupancy model (3.9 km) was about half the width of a gird cell. As we used grid-cell centroids to calculate distances between cells, our approach may not have allowed for smaller resolution of transmission distance (relative to the phylogenetic model) even though the rate of spread on the landscape was slower in the occupancy model. Although local transmission explained the majority of the occupancy probability, seasonality effects were also substantial (41% of explained variation). We found initial colonization probability to be approximately twice as high in the spring/summer months relative to the fall winter months, which is consistent with other studies of rabies seasonal prevalence in skunks [16].

Some previous studies of spatial spread of RABV in raccoons showed that landscape barriers such as major rivers, highways or mountains influence the direction and speed of spread [45,50], while other studies have not observed a strong effect of rivers and major roads [47,50]. In certain landscapes, the presence of suitable host habitat was observed to influence spread [51] and presence [52] of skunk RABV. Another study on spatial spread of skunk RABV in the mid-western USA observed differential impacts of river barriers depending on the skunk RABV variant [53]. Most of our study area was comprised of relatively flat terrain east of the Rocky Mountains. There were only two samples on the west side in the Rocky Mountains (none above 2509 meters), and our study area did not include the other side of the Rocky mountains, thus our models were unable to test effects of high-altitude barriers explicitly. In addition, there were no major rivers (e.g., Colorado River) in our study area, but there were two moderately sizes rivers (Poudre River and Big Thompson River) and one major road (Interstate 25). Although the rivers and roads could have been easily incorporated into our occupancy framework, we did not include them because: 1) the phylogenetic analysis suggested minimal heterogeneity in rates of spatial spread across the rivers and roads in our study

area, and 2) our dataset was small relative to the number of parameters we wanted to estimate (i.e., we prioritized quantification of spatial parameters). Also, maps of the genetic data showed closely related rabies variants on both sides of the most major road (Interstate 25) and rivers (Poudre River, Big Thompson River) throughout the study. Thus, although we did not explicitly test for effects of rivers and roads with our occupancy model, plots of the genetic data and the phylogenetic analyses suggested that the rivers and roads in our study area may not have presented major barriers that could reinforce controls against geographic spread [54]. If an intervention were deemed to be a necessary and/or cost-effective response, the primarily local transmission suggested that potential interventions could be applied as a relatively thin barrier (10–20 km) at the wave front to curb invasion, which is less than half the ~40km width typically used during aerial vaccination operation against raccoon rabies in the USA (Rich Chipman, National Rabies Management Program, personal communication). The strong seasonal difference in initial colonization probability suggests that interventions in the spring may be more effective than interventions in the fall.

Our model predicted that geographic spread tended to be mostly local and north-south, resulting in a wave-like pattern of geographic invasion. The phylogeographic analyses suggested that the invading viruses were new to Northern Colorado, such that the skunk population in Northern Colorado had not seen this virus in the recent past and was likely to be highly susceptible to invasion. We hypothesize that the reason we found a stronger tendency for movement of rabies from north to south instead of east to west, may be due to the underlying distribution of the skunk population. In our study area (called the Colorado Front Range), the human population forms a north-south belt, sandwiched by lower density regions to the west (in the mountains) and the east (in the plains). Although the literature does not suggest a consistent relationship of increasing skunk density across the rural to urban gradient [55, 56], it is possible that skunk abundance was higher running north-south than east-west, which could partly explain our directional effects.

Forecasting spatio-temporal disease spread in any system is challenging, but especially so for wildlife diseases due to an often poor understanding of host population abundance and ecology. Occupancy models present an alternative to approaches that rely on host demographic data, while quantifying rates and patterns of geographic spread based on underlying ecological processes. Although we accounted for observation error (by modeling the effect of human population size on prevalence), we did not have detailed data informing the observation process. For example, the study area included many natural areas spanning large regions where there are no human residents, but potentially heavy public usage, year-round. Accounting for reporting bias through collection of system-specific appropriate metadata (e.g. temporal usage patterns of the landscape by humans, geographic differences in the likelihood of human reporting) could improve geographic spread predictions using occupancy modeling. Our study showed how quantifying spatial parameters governing geographic spread can provide important ecological insights for understanding spatial epidemiology of rabies, conducting risk assessment and planning interventions—contributing to the toolbox of approaches for guiding management of disease invasions [57].

Supporting information

S1 Text. Methods. Additional tables describing surveillance data, a schematic describing the modeling strategy, and additional description of model formulation, validation, selection, and goodness of fit methodology.
(PDF)

S1 Figures. Results. Additional figures showing parameter fits and spatial spread. (PDF)

S1 Data. (CSV)

Acknowledgments

Author Contributions

Conceptualization: Kim M. Pepin, Amy J. Davis, Daniel G. Streicker, Amy T. Gilbert.

Data curation: Amy T. Gilbert.

Formal analysis: Kim M. Pepin, Amy J. Davis, Daniel G. Streicker, Justin W. Fischer.

Funding acquisition: Amy T. Gilbert.

Investigation: Kim M. Pepin, Daniel G. Streicker.

Methodology: Kim M. Pepin, Amy J. Davis, Daniel G. Streicker, Amy T. Gilbert.

Project administration: Kim M. Pepin, Amy T. Gilbert.

Resources: Kim M. Pepin, Kurt C. VerCauteren, Amy T. Gilbert.

Software: Kim M. Pepin.

Supervision: Kim M. Pepin, Amy T. Gilbert.

Validation: Kim M. Pepin, Amy J. Davis, Daniel G. Streicker.

Visualization: Amy T. Gilbert.

Writing – original draft: Kim M. Pepin.

Writing – review & editing: Kim M. Pepin, Amy J. Davis, Daniel G. Streicker, Justin W. Fischer, Kurt C. VerCauteren, Amy T. Gilbert.

References

1. McClintock BT, Nichols JD, Bailey LL, MacKenzie DI, Kendall W, et al. (2010) Seeking a second opinion: uncertainty in disease ecology. Ecol Lett 13: 659–674. https://doi.org/10.1111/j.1461-0248.2010.01472.x PMID: 20426794

2. Blanton JD, Palmer D, Rupprecht CE (2010) Rabies surveillance in the United States during 2009. JAVMA 237: 646–657. https://doi.org/10.2460/javma.237.6.646 PMID: 20839985

3. Buenconsejo J, Fish D, Childs JE, Holford TR (2008) A Bayesian hierarchical model for the estimation of two incomplete surveillance data sets. Stat Med 27: 3269–3285. https://doi.org/10.1002/sim.3190 PMID: 18314934

4. Koffi JK, Leighton PA, Pelcat Y, Trudel L, Lindsay LR, et al. (2012) Passive surveillance for I. scapularis ticks: enhanced analysis for early detection of emerging Lyme disease risk. J Med Entomol 49: 400–409. PMID: 22493860

5. Selemani M, Mrema S, Shamte A, Shabani J, Mahande MJ, et al. (2015) Spatial and space—time clustering of mortality due to malaria in rural Tanzania: evidence from Ifakara and Rufiji Health and Demographic Surveillance System sites. Malaria J 14: 1.

6. Kim BI, Blanton JD, Gilbert A, Castrodale L, Hueffer K, et al. (2014) A conceptual model for the impact of climate change on fox rabies in Alaska, 1980–2010. Zoonoses Public Hlth 61: 72–80.

7. Rupprecht CE, Hanlon CA, Hemachudha T (2002) Rabies re-examined. Lancet Infect Dis. 2: 327–343. PMID: 12144896

8. Hampson K, Coudeville L, Lembo T, Sambo M, Kieffer A, et al. (2015) Estimating the global burden of endemic canine rabies. PLoS Negl Trop Dis 9: e0003709. https://doi.org/10.1371/journal.pntd.0003709 PMID: 25881058

9. Favoretto S, De Mattos C, De Mattos C, Campos A, Sacramento D, et al. (2013) The emergence of wildlife species as a source of human rabies infection in Brazil. Epid Infect 141: 1552–1561.

10. Velasco-Villa A, Reeder SA, Orciari LA, Yager PA, Franka R, et al. (2008) Enzootic rabies elimination from dogs and reemergence in wild terrestrial carnivores, United States. Emerg Infect Dis 14: 1849. https://doi.org/10.3201/eid1412.080876 PMID: 19046506

11. Chiou H-Y, Hsieh C-H, Jeng C-R, Chan F-T, Wang H-Y, et al. (2014) Molecular characterization of cryptically circulating rabies virus from ferret badgers, Taiwan. Emerg Infect Dis 20: 790. https://doi.org/10.3201/eid2005.131389 PMID: 24751120

12. Chipman R, Slate D, Rupprecht C, Mendoza M (2008) Downside risk of wildlife translocation. Development in Biologicals 131: 223–232.

13. Rupprecht CE, Smith JS, Fekadu M, Childs JE (1995) The ascension of wildlife rabies: a cause for public health concern or intervention?. Emerg Infect Dis 1: 107. https://doi.org/10.3201/eid0104.950401 PMID: 8903179

14. Wallace RM, Gilbert A, Slate D, Chipman R, Singh A, et al. (2014) Right place, wrong species: a 20-year review of rabies virus cross species transmission among terrestrial mammals in the United States. PLoS O 9: e107539.

15. Slate D, Algeo TP, Nelson KM, Chipman RB, Donovan D, et al. (2009) Oral rabies vaccination in North America: opportunities, complexities, and challenges. PLoS Negl Trop Dis 3: e549. https://doi.org/10.1371/journal.pntd.0000549 PMID: 20027214

16. Oertli EH, Wilson PJ, Hunt PR, Sidwa TJ, Rohde RE (2009) Epidemiology of rabies in skunks in Texas. JAVMA 234: 616–620. https://doi.org/10.2460/javma.234.5.616 PMID: 19250039

17. Smith JS, Reid-Sanden FL, Roumillat LF, Trimarchi C, Clark K, et al. (1986) Demonstration of antigenic variation among rabies virus isolates by using monoclonal antibodies to nucleocapsid proteins. J Clin Microbiol 24(4): 573–580. PMID: 2429983

18. Kuzmina NA, Lemey P, Kuzmin IV, Mayes BC, Ellison JA, et al. (2013) The phylogeography and spatiotemporal spread of south-central skunk rabies virus. PLoS One 8: e82348. https://doi.org/10.1371/journal.pone.0082348 PMID: 24312657

19. Gilbert A, Kohler D, Rigg T, Fischer J, Spraker T, et al. (2014) A recent epizootic of skunk rabies and associated spillover in Northern Colorado, USA. Proceedings of the 26th Vertebrate Pest Conference. 2014; pp 323–327.

20. MacKenzie DI, Nichols JD, Royle JA, Pollock K, Bailey L, et al. (2006) Occupancy estimation and modelling—inferring patterns and dynamics of species occurrence. 2006; San Diego, CA, Elsevier.

21. Bailey LL, MacKenzie DI, Nichols JD (2014) Advances and applications of occupancy models. Meth Ecol Evol 5: 1269–1279.

22. Miller DA, Talley BL, Lips KR, Campbell Grant EH (2012) Estimating patterns and drivers of infection prevalence and intensity when detection is imperfect and sampling error occurs. Meth Ecol Evol 3: 850–859.

23. Walter KS, Pepin KM, Webb CT, Gaff HD, Krause PJ, et al. (2016) Invasion of two tick-borne diseases across New England: harnessing human surveillance data to capture underlying ecological invasion processes. Proc Roy Soc B 283: 20160834.

24. Bright EA, A.N. R (2013) LandScan Oak Ridge. In: Oak Ridge National Laboratory, editor: http://www.ornl.gov/sci/landscan.

25. Charlton KM, Webster WA, Casey GA (1991) Skunk rabies. In: Baer GM, editor. *The natural histories of rabies*, 2nd edition. Boca Raton, FL, USA: CRC Press, Inc..

26. Guerra MA, Curns AT, Rupprecht CE, Hanlon CA, Krebs JW, et al. (2003) Skunk and raccoon rabies in the eastern United States: temporal and spatial analysis—research. Emerg Infect Dis 9: 1143–1150. https://doi.org/10.3201/eid0909.020608 PMID: 14519253

27. Bixler A, Gittleman JL (2000) Variation in home range and use of habitat in the striped skunk (*Mephitis mephitis*). J Zool 251: 525–533.

28. Weissinger MD, Theimer TC, Bergman DL, Deliberto TJ (2009) Nightly and seasonal movements, seasonal home range, and focal location photo-monitoring of urban striped skunks (*Mephitis mephitis*): implications for rabies transmission. J Wild Dis 45: 388–397.

29. Bled F, Royle JA, Cam E (2011) Hierarchical modeling of an invasive spread: the Eurasian Collared-Dove Streptopelia decaocto in the United States. Ecol Appl 21: 290–302. PMID: 21516906

30. Gelman A, Carlin JB, Stern HS, Dunson DB, Vehtari A, et al. (2013) Bayesian Data Analysis, Third Edition. Boca Raton, FL, USA: Taylor & Francis.

31. R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.

32. Hooten M, Hobbs N (2015) A guide to Bayesian model selection for ecologists. Ecol Monog 85: 3–28.

33. Watanabe S (2013) A widely applicable Bayesian information criterion. J Mach Learn Res 14: 867–897.

34. Fawcett T (2006) An introduction to ROC analysis. Patt Rec Lett 27: 861–874.

35. Long RA, Donovan TM, MacKay P, Zielinski WJ, Buzas JS (2011) Predicting carnivore occurrence with noninvasive surveys and occupancy modeling. Land Ecol 26: 327–340.

36. Burnham KP, Anderson DR (2004) Multimodel inference understanding AIC and BIC in model selection. Sociol Meth Res 33: 261–304.

37. Powell LA (2007) Approximating variance of demographic parameters using the Delta Method: A reference for avian biologists. Condor 109: 949–954.

38. Davis R, Nadin-Davis SA, Moore M, Hanlon C (2013) Genetic characterization and phylogenetic analysis of skunk-associated rabies viruses in North America with special emphasis on the central plains. Virus Res 174: 27–36. https://doi.org/10.1016/j.virusres.2013.02.008 PMID: 23524137

39. Drummond AJ, Suchard MA, Xie D, Rambaut A (2012) Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol Biol Evol 29: 1969–1973. https://doi.org/10.1093/molbev/mss075 PMID: 22367748

40. Rambaut A, Lam TT, Max Carvalho L, Pybus OG (2016) Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2: vew007. https://doi.org/10.1093/ve/vew007 PMID: 27774300

41. Baele G, Lemey P, Bedford T, Rambaut A, Suchard MA, et al. (2012) Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol Biol Evol 29: 2157–2167. https://doi.org/10.1093/molbev/mss084 PMID: 22403239

42. Darriba D, Taboada GL, Doallo R, Posada D (2012) jModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9: 772.

43. Dellicour S, Rose R, Pybus OG (2016) Explaining the geographic spread of emerging epidemics: a framework for comparing viral phylogenies and environmental landscape data. BMC Bioinformatics 17: 82. https://doi.org/10.1186/s12859-016-0924-x PMID: 26864798

44. Wesolowski A, Qureshi T, Boni MF, Sundsøy PR, Johansson MA, et al. (2015) Impact of human mobility on the emergence of dengue epidemics in Pakistan. PNAS 112: 11887–11892. https://doi.org/10.1073/pnas.1504964112 PMID: 26351662

45. Smith DL, Lucey B, Waller LA, Childs JE, Real LA (2002) Predicting the spatial dynamics of rabies epidemics on heterogeneous landscapes. PNAS 99: 3668–3672. https://doi.org/10.1073/pnas.042400799 PMID: 11904426

46. Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. PNAS 104: 7993–7998. https://doi.org/10.1073/pnas.0700741104 PMID: 17470818

47. Rioux Paquette S, Talbot B, Garant D, Mainguy J, Pelletier F (2014) Modelling the dispersal of the two main hosts of the raccoon rabies variant in heterogeneous environments with landscape genetics. Evol App 7: 734–749.

48. Hampson K, Dushoff J, Cleaveland S, Haydon DT, Kaare M, et al. (2009) Transmission dynamics and prospects for the elimination of canine rabies. PLoS Biol 7: e1000053.

49. Mollentze N, Nel LH, Townsend S, Le Roux K, Hampson K, et al. (2014) A Bayesian approach for inferring the dynamics of partially observed endemic infectious diseases from space-time-genetic data. Proc Roy Soc B 281: 20133251.

50. Biek R, Henderson JC, Waller LA, Rupprecht CE, Real LA (2007) A high-resolution genetic signature of demographic and spatial expansion in epizootic rabies virus. PNAS 104: 7993–7998. https://doi.org/10.1073/pnas.0700741104 PMID: 17470818

51. Ramey CA, Mills KW, Fischer JW, McLean RG, Fagerstone KA, et al. (2013) Graphically characterizing the movement of a rabid striped skunk epizootic across the landscape in Northwestern Wyoming. EcoHealth 10: 246–256. https://doi.org/10.1007/s10393-013-0853-3 PMID: 23812724

52. Raghavan RK, Hanlon CA, Goodin DG, Davis R, Moore M, et al. (2016) Bayesian spatiotemporal pattern and eco-climatological drivers of striped skunk rabies in the North Central Plains. PLoS Negl Trop Dis 10: e0004632. https://doi.org/10.1371/journal.pntd.0004632 PMID: 27127994

53. Barton HD, Gregory AJ, Davis R, Hanlon CA, Wisely SM (2010) Contrasting landscape epidemiology of two sympatric rabies virus strains. Mol Ecol 19: 2725–2738. https://doi.org/10.1111/j.1365-294X.2010.04668.x PMID: 20546130

54. Real LA, Biek R (2007) Spatial dynamics and genetics of infectious diseases on heterogeneous landscapes. J Roy Soc Interface 4: 935–948.

55. Helgen K and Reid F (2016) Mephitis mephitis. The IUCN red list of threatened species 2016: e.T41635A45211301. http://dx.doi.org/10.2305/IUCN.UK.2016-1.RLTS.T41635A45211301.en.

56. Rosatte R, Sobey K, Dragoo JW, Gehrt SD. Striped skunks and allies. In: Gehrt SD, Riley SP, Cypher BL, editors. Urban carnivores: ecology, conflict, and conservation. JHU Press; 2010. pp. 97–108.

57. Dicko AH, Lancelot R, Seck MT, Guerrini L, Sall B, Lo M, Vreysen MJB, Lefrancois T, Fonta WM, Peck S, Bouyer J. 2014. Using species distribution models to optimize vector control in the framework of the tsetse eradication campaign in Senegal. PNAS 111:10149–10154. https://doi.org/10.1073/pnas.1407773111 PMID: 24982143