

strategy. Independent Bernoulli prior probabilities are assigned to each δ_n while independent normal distributions are used as priors for β_n . We calculate Bayes factors (BF) of the inclusion parameters (posterior odds divided by the corresponding prior odds) to identify the factors that have statistically-supported impacts on migration ($BF > 3$)⁷⁶. Model selection and parametrization are implemented in BEAST 2.0 where the joint posterior probability and marginalizations are approximated using standard Markov chain Monte Carlo (MCMC). This modeling approach accounts for (a) uncertainty in ancestral state reconstruction; (b) highly-correlated predictors through multivariate effect-size parameters (β^*); and (c) potential sampling biases by including sample sizes as diffusion predictors^{76,155}. This powerful modeling approach allows us to explore the effect of all possible environmental factors and combinations of factors simultaneously and provide the marginal degree of support for each factor (*i.e.*, is there statistical support suggesting that tick density affects emigration rates?) and the degree and direction of correlation (*i.e.*, to what extent does tick density increase or decrease emigration rates?)^{18,68,70,156–158}. This framework identifies factors associated with sampling sites, as well as coarse-scale inter-site factors, that expedite or impede emigration from colonized locations and *imm*migration to colonized sites or uncolonized locations. In **Aim 3**, the statistically supported environmental factors identified here will be used as variables in geospatial models to estimate future geographic spread of each pathogen and the accuracy of these estimates will be assessed using prospective sampling.

Landscape resistance: Landscape genetic approaches quantify the impact on migration of fine-scale environmental features between sampling sites, areas that are considered only at coarse scales in the discrete-site phylogeographic diffusion model framework^{131,159,160}. We follow the highly-effective iterative framework previously described that first uses linear regression models fitted with generalized least squares to estimate the impact of each environmental feature (predictor variables, **Table 3** and **4**) on the genetic distance between samples (response variable)¹⁶¹. Model selection is accomplished using AICc and the variance explained by the fixed effects and quantile-quantile plots of residuals of each model are used to assess the goodness-of-fit of each model^{161,162}. Variance in model residuals, which tend to increase with genetic distance in our experience, are fitted with power functions when necessary. Cross-validation will be accomplished by fitting the model to a randomly chosen 75% of the data and predicting the remaining 25% of the data from this model. This process will be repeated 100 times and the distribution in prediction error will be used as an indicator to compare the 10 best (AICc) models¹⁶¹. Model results can be represented as maps of resistance to migration where each environmental feature in a landscape has a quantifiable impact on pathogen dispersal^{149,163,164}. Resistance maps will initially be built using the AICc-best model parameters and Euclidian distance and will be modified by iteratively calculating least-cost paths, identifying improved submodels, and re-building resistance surfaces¹⁶¹. Briefly, the least-cost paths that explain genetic distances will first be built from direct paths on the initial map and the least-cost distances calculated using *raster* and *gdistance*^{165,166}. The averages of all environmental variables will then be recalculated along a least-cost path and the Euclidian distances will be replaced by the new least-cost distances. This process allows refitting of the model to the data in order to select an improved submodel by AICc. The landscape will then be repopulated with the revised resistance map. This iterative process will continue until recalculated models do not improve AICc (generally 7-10 iterations). This simple process -

Table 3 – Environmental variables

Dataset	Type	Source	Scale
Climate (mean and coldest ¼ temp, precipitation)	raster	Bioclim ¹³⁵	30"
Digital Elevation Model (DEM)	raster	SRTM ¹³⁶	90m
Lakes and rivers	vector	ESRI ¹³⁷	N/A
Land cover (forest and soil type/cultivated land/urban)	raster	USGS/USDA ^{138,139,141,142}	30m
Vertebrate communities	raster	GAP/survey ^{145–147}	30m
Human density	vector	TIGER ¹⁴³	N/A
Road traffic density	vector	NYDoT ¹⁴⁰	N/A

Table 4 – Derived environmental variables¹⁴⁴

Dataset	Type	Derived from:
Aspect	raster	DEM
Distance to developed areas	raster	land cover
Distance to roads	raster	roads map
Land use change	raster	land cover
Slope	raster	DEM