features associated with those patterns[15,18,59,65–70]. The research team, with relevant expertise in tick-borne pathogen biology (Brisson/NYDoH), genome sequencing and analyses (Brisson), phylodynamics and geospatial modeling (Brisson), standardized field sampling (NYDoH), and public health translation (NYDoH), has collaborated since 2011 on this disease system. Our prior studies (a) identified, and quantified the effect size of, factors affecting dynamic patterns of *I. scapularis* populations; (b) used these data to parameterize demographic models; (c) estimated future range and population size increases from the demographic models; and (d) assessed the accuracy of these estimates[15,18,21]. These studies produced consequential insights with relevance to public health and ecological, evolutionary, and biological aspects of this important disease system. These studies, combined with other projects and the current literature, form the rationale of the proposed work and have aided in designing protocols that address deficiencies in prior studies. The samples (75,887 ticks collected from 515 sites) from NY State - which is representative of much of the natural environment that ticks encounter in the northeastern US including rapid and recent changes in climate and landscapes - provide the data necessary to quantify the effect of environmental features on pathogen demography and validate the resulting models across a broad geographic area (**Fig. 2**).

Preliminary data

C.1 - Emprical validations. Assessing the historical, ecological, landscape, and human factors that affect heterogeneity in population demography over both time and space is a fundamental scientific problem, especially for disease-causing systems[1–4,19,29,67,71–74]. Large scale features associated with heterogeneity in tick demography in the Hudson Valley were assessed using temporally- and spatially-structured field collected data (2003-2008) in a geospatial modeling framework[15]. These analyses identified environmental and climatic factors that accounted for the observed spatial and temporal dynamics. The models were then used to estimate future tick dynamics within the Hudson Valley and across NY State[21]. Our large-scale sampling permitted empirical assessment of model estimates, a feat rarely possible for evolutionary or ecological models (**Fig. 2**). These models accurately forecasted tick densities within the Hudson River Valley (R^2=0.69) and across all sampling sites in NY state (R^2=0.72). The accuracy of the model estimates across a broad and diverse section of the northeastern US demonstrates their potential to accurately estimate future human disease risk, a crucial component of public health.

C.2 - Phylodynamic diffusion modeling. The historical and migratory factors that influenced heterogeneity in tick demography were assessed in a phylodynamic diffusion modeling framework (**Box 1**). These analyses demonstrated a recent northward geographic range expansion in which uncolonized areas were colonized by migrants from nearby southern locations (**Fig. 3**). Newly colonized areas increased in population size and became the source of migrants to proximal areas to the north[18]. The influence of environmental factors on migration rates were assessed by parameterizing the pairwise diffusion rates as a function of potential environmental predictors including geographic distance, climate, land cover[75,76]. These analyses found statistical support for coarse-scale factors similar to those identified in our previous analyses including geographic distance, precipitation, winter temperature, and forest cover[15,75]. These data, as well as the pathogen genome data, meet all assumptions of these methods for quantifying impacts of environmental factors on the spatial diffusion process (**Aim 2**).

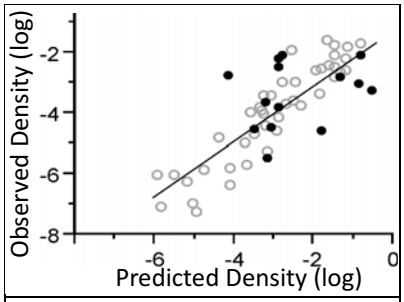

Fig 2. **Predictive models built using Hudson River Valley data (2003-8) accurately estimated future tick densities across New York State (2012-13).** Demographic models accurately estimated tick densities within the Hudson Valley (solid points; R^2=0.69) and across all NY state sites (open circles; R^2=0.72).

C.3 - Selective Whole Genome Amplification (SWGA). SWGA is an exciting innovation we developed to enable microbial population genomic studies[49,65]. Commercial whole genome amplification methods use the ϕ29 polymerase primed with random hexamers to amplify large DNA fragments (~70kb) with excellent accuracy (two-orders of magnitude less error-prone