

RESEARCH

Open Access



Phylogeographic dynamics of the arthropod vector, the blacklegged tick (*Ixodes scapularis*)

Kayleigh R. O’Keeffe^{1*}, Zachary J. Oppler¹, Melissa Prusinski², Richard C. Falco², JoAnne Oliver³, Jamie Haight², Lee Ann Sporn⁴, P. Bryon Backenson² and Dustin Brisson¹

Abstract

Background: The emergence of vector-borne pathogens in novel geographic areas is regulated by the migration of their arthropod vectors. Blacklegged ticks (*Ixodes scapularis*) and the pathogens they vector, including the causative agents of Lyme disease, babesiosis and anaplasmosis, continue to grow in their population sizes and to expand in geographic range. Migration of this vector over the previous decades has been implicated as the cause of the re-emergence of the most prevalent infectious diseases in North America.

Methods: We systematically collected ticks from across New York State (hereafter referred to as New York) from 2004 to 2017 as part of routine tick-borne pathogen surveillance in the state. This time frame corresponds with an increase in range and incidence of tick-borne diseases within New York. We randomly sampled ticks from this collection to explore the evolutionary history and population dynamics of *I. scapularis*. We sequenced the mitochondrial genomes of each tick to characterize their current and historical spatial genetic structure and population growth using phylogeographic methods.

Results: We sequenced whole mitochondrial genomes from 277 ticks collected across New York between 2004 and 2017. We found evidence of population genetic structure at a broad geographic scale due to differences in the relative abundance, but not the composition, of haplotypes among sampled ticks. Ticks were often most closely related to ticks from the same and nearby collection sites. The data indicate that both short- and long-range migration events shape the population dynamics of blacklegged ticks in New York.

Conclusions: We detailed the population dynamics of the blacklegged tick (*Ixodes scapularis*) in New York during a time frame in which tick-borne diseases were increasing in range and incidence. Migration of ticks occurred at both coarse and fine scales in the recent past despite evidence of limits to gene flow. Past and current tick population dynamics have implications for further range expansion as habitat suitability for ticks changes due to global climate change. Analyses of mitochondrial genome sequencing data will expound upon previously identified drivers of tick presence and abundance as well as identify additional drivers. These data provide a foundation on which to generate testable hypotheses on the drivers of tick population dynamics occurring at finer scales.

Keywords: Ticks, Vectors, Lyme disease, Phylogeography, Migration, Population structure

Background

Vector-borne diseases are the most common type of emerging and re-emerging infectious diseases (EIDs) in the world and constitute major threats to public health [1, 2]. The recent surge in vector-borne diseases is a consequence of the effects of global climate change and human land-use change on arthropod vector population

*Correspondence: kokeeffe@sas.upenn.edu

¹ Department of Biology, University of Pennsylvania, Philadelphia, PA, USA
Full list of author information is available at the end of the article



© The Author(s) 2022. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

dynamics [3–9]. The distribution and abundance of vectors are major drivers in vector-borne disease spread [10, 11]. Therefore, identifying the ecological processes governing arthropod vector population dynamics is essential for predicting future spread of diseases caused by vector-borne pathogens [12]. Such investigations are challenging as they require a wealth of fine-scale temporal and spatial data that encompass a time frame in which demographic processes occurred [13–15]. Here, we address this challenge using samples of blacklegged ticks (*Ixodes scapularis*) collected across New York State (hereafter referred to as New York) during a time frame of demographic dynamics to delineate the patterns of tick population dynamics over time and space.

Ixodes scapularis is the most important vector of EIDs in the United States [16], with this tick vectoring pathogens that cause numerous diseases, such as Lyme disease, human babesiosis and anaplasmosis [17]. The population density and geographic range of *I. scapularis* have increased in the northeastern United States over recent decades [10, 18–21]. For example, the range of ticks in New York has expanded markedly into the western and northern parts of the state during the last decade alone [22].

Tick population dynamics are a primary driver of the population dynamics of tick-transmitted pathogens. For example, changes in the population demography of ticks in the Hudson River Valley explained over 40% of changes in human Lyme disease incidence in the same area and time frame [12]. Investigating tick population dynamics across broad geographic and temporal ranges with sufficient analytical power to identify the ecological processes underlying changes in tick populations is critical to the design of effective control strategies [23–30]. However, the majority of studies have been limited in geographic and temporal range due to the difficulty of obtaining samples at the appropriate temporal and spatial scales [10, 12]. Exploratory analyses that delineate tick population dynamics at coarse scales provide the foundation to generate hypotheses about the fine-scale ecological processes that influence tick and tick-borne pathogen populations. Subsequent analyses of targeted samples collected at finer scales are necessary to address these hypotheses.

Exploratory research has been the cornerstone for hypothesis generation across scientific fields and is key to furthering our understanding of vector biology. For example, genome-wide association studies (GWAS) have facilitated the generation of numerous hypotheses about the function of multiple genes in traits and diseases, which have then been tested in subsequent functional investigations [31–33]. Exploratory research, such as GWAS, that facilitates broad-scale inferences and

hypothesis generation are feasible due to the increase in the availability of massive datasets on genomes, species distributions, climate, land cover and other types of 'big data' [34, 35]. By taking a similar exploratory approach with an extensive sample collection of ticks, we characterized natural vector population dynamics at broad spatial scales to generate hypotheses on how biological, environmental and human features impact such dynamics.

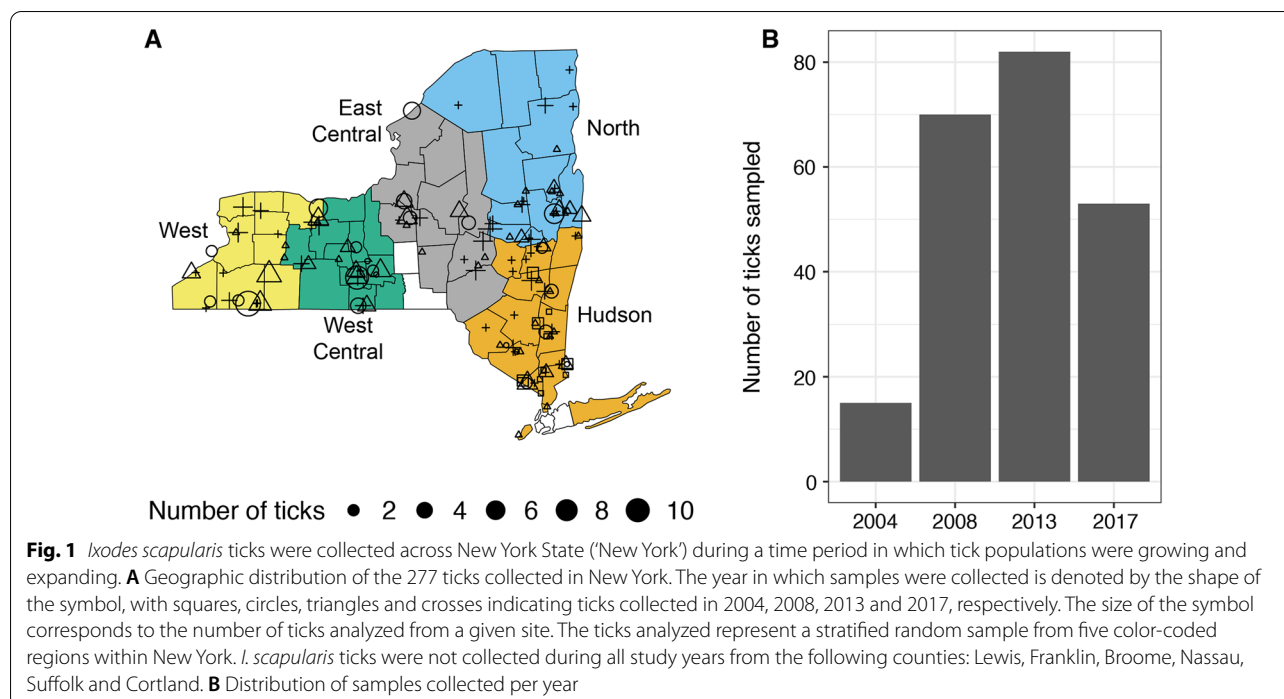
In this study, we describe the population dynamic patterns of blacklegged ticks in New York using a sample collection that represents a time frame and geographic range in which tick populations and tick-borne disease incidence has grown and expanded. Specifically, we analyzed the mitochondrial genomes from 277 *I. scapularis* ticks collected across the state of New York from 2004–2017. We inferred tick population dynamics and generated hypotheses about the mechanisms underlying those dynamics using the additional statistical power derived from a sample selection representing state-wide environmental variation and abundant genetic information from each tick. Specifically, we (1) characterized population genetic structure and (2) characterized population growth. This work is critical to designing effective control and management strategies for ticks and tick-borne diseases.

Methods

Population sampling

Host-seeking ticks have been collected since 2003 using standardized flagging methods [36] as part of routine tick-borne pathogen surveillance in New York [37]. This collection includes *I. scapularis* ticks and their associated pathogens from 667 sites [22, 37]. Some locations were sampled annually while others were sampled on a rotational basis every 2 to 5 years. Collections have been systematic with respect to both space and time. Collected ticks were stored in 80–100% ethanol at 4 °C until sorted by developmental stage and confirmed to be *I. scapularis* using dichotomous keys [38, 39]. Following initial processing, individual ticks were placed in 1.5-ml Eppendorf tubes and preserved in 80–100% ethanol at – 20 °C until DNA extraction.

We randomly sampled 400 ticks from this collection to investigate current and historical patterns of population growth and gene flow across New York. We employed a stratified random sample design to maximize spatial coverage while minimizing violations of the assumptions of the subsequent analyses. We divided New York into five approximately equally-sized regions (Fig. 1a; these regions are referred to in this article as Hudson/Hudson River Valley, North, East Central, West Central and West). From each region, we randomly sampled



ticks collected in each of three different years (2008, 2013 and 2017). Specifically, 25 ticks were sampled from each region and year combination. In addition, 25 ticks were sampled from those collected in 2004 from the Hudson River Valley, the only region in which ticks were collected prior to the initiation of state-wide surveillance in 2008. At least one tick was randomly sampled from each county when available to maximize the spatial distribution of the samples within each region. Sample metadata, including sampling year, location, region, developmental stage, and sex, are listed in Additional file 1: Table S1.

Several ticks from outside of New York were included to serve as geographic outgroups. Specifically, we included three ticks collected from southeastern Pennsylvania (2 collected in 2009, 1 in 2010 from Crow's Nest Preserve [Natural Lands Trust], a deciduous mid-Atlantic forest with interspersed agricultural lands [40]) in Chester County (40°11' N, 75°17' W), and three ticks collected from Loudoun County, Virginia in 2017 (collected by Jory Brinkerhoff at the University of Richmond).

DNA extractions

Total genomic DNA was extracted from individual *I. scapularis* ticks as previously described [37]. Briefly, ticks were double rinsed with nuclease-free distilled water and placed in 2.0-ml round-bottomed Safe-Lock tubes containing two 4-mm stainless steel grinding beads and 180 μ l phosphate buffered saline (PBS), pH 7.2. Samples were ground using an electric tissue homogenizer. Whole

genomic DNA was extracted from homogenized samples using the DNeasy™ Blood and Tissue kit (Qiagen, Hilden, Germany) following the manufacturer's recommendations (Purification of Total DNA from Insects protocol). DNA from individual ticks was initially eluted in 200 μ l ultra-pure water and stored at -20°C . In the case of the 2017 samples, second elutions of 100 μ l ultra-pure water were used for sequencing.

Mitochondrial genome amplification and sequencing

We designed primers for long-range PCR to capture the entire mitochondrial genome of each tick. Briefly, we developed primers that amplify the mitochondrial genome in four separate but overlapping fragments (Additional file 2). We used the *I. ricinus* reference genome as the basis for primer design with the NCBI primer design tool [41]. We chose primer pairs that fell within regions identified as highly conserved and slightly overlapped one another. Primer sequences are reported in Additional file 2: Table S2. Each fragment was amplified separately using the same PCR program. Each 50- μ l PCR reaction volume contained 1 μ l GXL polymerase (TaKaRa Bio, Kusatsu, Shiga, Japan), 10 μ l of 5 \times GXL PCR buffer, 4 μ l GXL dNTPs, 5 μ l forward primer, 5 μ l reverse primer, 15 μ l water and 10 μ l DNA. The cycling conditions were: initial denaturation at 95 $^{\circ}\text{C}$ for 30 s; followed by 5 cycles of 98 $^{\circ}\text{C}$ for 10 s, 64 $^{\circ}\text{C}$ for 15 s, 68 $^{\circ}\text{C}$ for 4.5 min, 5 cycles of 98 $^{\circ}\text{C}$ for 10 s, 67 $^{\circ}\text{C}$ for 15 s, 68 $^{\circ}\text{C}$ for 4.5 min and 30 cycles of 98 $^{\circ}\text{C}$ for 10 s, 70 $^{\circ}\text{C}$ for 15 s,

68 °C for 4.5 min. PCR product quality was determined by visualizing products using gel electrophoresis and quantifying concentration using Qubit. PCR product concentrations were normalized to 10 ng/μl and pooled in one run on an Illumina MiSeq instrument (Illumina, San Diego, CA, USA) using a paired-end 2 × 300 bp kit (Illumina).

Read alignment and SNP detection

Reads were demultiplexed following Illumina's Generate FASTQ workflow and trimmed with CutAdapt (version 1.18 [42]). We aligned reads to both the *I. ricinus* [43] and *Ixodes scapularis* [44] mitochondrial genomes using BWA mem (v. 0.7.17 [45, 46]). The resulting set of variant calling files were 99.2% identical, and there was no difference in between results of subsequent analyses. Duplicate reads were excluded from downstream analysis using the Picard Suite (v. 2.18.16 [47]) MarkDuplicates. Only samples with mean coverage > 10× were retained for analysis. Single nucleotides that differed from the draft genome were identified with GATK (v. 4.2.0 [48, 49]) HaplotypeCaller (ploidy set to 1) to generate sample-specific gVCF files. We conducted joint genotyping of samples with GATK GenotypeGVCFs. We excluded indels and single nucleotide polymorphisms (SNPs) with signals of low mapping or genotyping quality with GATK VariantFiltration using filters recommended by GATK: quality by depth (QD < 2.0), Fisher strand bias (FS > 60.0), mapping quality (MQ < 40.0), mapping quality rank sum test (MQRankSum < -12.5), the Mann–Whitney rank sum test (ReadPosRankSum < -8.0), low genotype call (GQ < 20) and strand odds ratio test (SOR > 4.0). Filtered genotypes were set to no call (GATK option: setFilteredGtToNocall). We used the GATK tool FastaAlternateReferenceMaker to create an alignment of consensus sequences from the variant file. We accounted for no calls within each sample in two ways: (i) in one set of consensus files, reference bases were used where there was no call made; (ii) in another set of consensus files, no calls were marked with *Ns*. FastaAlternateReferenceMaker outputs consensus files use reference bases where no calls are made by default. To mark no calls with *Ns*, we used SelectVariants to select sites in which no call could be made, and the resulting VCF file was used as a mask file (using the snp-mask parameter) when running FastaAlternateReferenceMaker. This resulted in two sets of alignments of consensus sequences of the mitochondria (14,420 bp). We ran all analyses with both sets of files, and found that our inferences were robust to the method used. We report the results with *Ns* in the main text.

Phylogeographic analyses

We used a coalescent approach implemented in BEAST 2 (v. 2.6.3 [50]) to infer the phylogeographic history of *I. scapularis*. BEAST estimates the phylogenetic topology while simultaneously inferring the geographic location of ancestral nodes, migratory events among locations and demography. Recombination-free alignments were used to fit two alternative models, including either a strict molecular clock or a relaxed (uncorrelated log-normal) molecular clock (allowing substitution rates to vary across tree branches). Each model included a Bayesian skyline model of changing population size, an Hasegawa-Kishino-Yano (HKY) substitution model and gamma-distributed rate variation across sites.

We ran chains of 500 million iterations or until convergence for each model. Chains were thinned by sampling every 50,000 iterations, and 10% of each chain was discarded as burn-in. Tracer v. 1.7.1 was used to assess convergence and confirm effective sample sizes were > 200 for each parameter [51]. We used path sampling to estimate the marginal likelihood of each model to identify the best-fitting model [52]. We compared alternative models with Bayes factors and identified the relaxed log-normal clock as the best-fitting model used in all above analyses. Maximum clade credibility (MCC) trees were generated using TreeAnnotator [53], and trees were visualized in FigTree (<http://tree.bio.ed.ac.uk/software/figtree/>).

We complemented the coalescent-based approach with a mismatch distribution analysis implemented in DnaSP v5 [54]. We tested if our observed data were consistent with a mismatch distribution expected from a sudden population expansion or a spatial expansion. The observed mismatch distribution and expected mismatch distribution for a stable population were visualized in the ggplot2 package (v. 3.3.3 [55]) in R v. 3.6.1.

Population demography

We quantified pairwise genetic distances among samples using the Kimura 2-parameter model [56] in MEGA X [57] to investigate genetic differentiation. Genetic differentiation was estimated both within regions and between regions by comparing pairwise genetic distances between ticks. We used linear models with R (version 3.6.1) to compare within-region pairwise distances across regions and over time (Distance ~ Region × Year) and between-region pairwise distances across region pairs (Distance ~ Region Pair). We used the package *emmeans* (version 1.3.2 [58]) to evaluate the estimated marginal means of diversity indices for each explanatory variable level, adjusted for multiple comparisons (Tukey's honest significant difference [HSD]).

To test for isolation by distance [59], we performed a Mantel permutation test [60] with 9999 permutations of geographic distance between collection sites and genetic distance as quantified by the Kimura 2-parameter model. The haversine distance between ticks was calculated from the longitude and latitude of tick collection sites using the *distm* function within the *vegan* library in R [61]. We performed the Mantel test using the *mantel* function within *vegan*. The Mantel statistic, r_{M^2} quantifies the correlation between genetic and geographic distance.

Results

Field sampling and variation

We sequenced whole mitochondrial genomes from 277 ticks collected in New York (Table 1; Fig. 1). The 187 identified SNPs resulted in 88 unique haplotypes among New York ticks (Table 1). The number of haplotypes and haplotype diversity were highest in the Hudson River Valley (Hudson region) and lowest in the Northern region of New York after correcting for differences in sample sizes. The six ticks from outside of New York (3 from Virginia, 3 from Pennsylvania) had nine additional SNPs, resulting in an additional two haplotypes.

Population genetic structure among ticks in New York

There was clear evidence of population genetic structure among the five regions delineated within New York.

Table 1 Genetic diversity indices including the number of haplotypes and haplotype diversity of *Ixodes scapularis* for each of the five regions sampled in New York State

Genetic diversity indices	Regions				
	Hudson	North	East Central	West Central	West
Sample size (n ticks)	79	35	44	51	62
H	48	22	28	33	40
H_d	0.983	0.965	0.970	0.966	0.981

H Number of haplotypes, H_d haplotype diversity

Table 2 Population genetic structure of *Ixodes scapularis* among sampled regions of New York State

Regions ^a	Regions ^a				
	Hudson (48)	North (22)	East Central (28)	West Central (33)	West (40)
Hudson (48)	–	0.016	0.011	0.009	0.007
North (22)	11/59	–	0.014	0.017	0.016
East Central (28)	13/63	5/45	–	0.009	0.012
West Central (33)	16/65	9/46	13/48	–	0.006
West (40)	12/76	7/55	11/57	17/56	–

^a Pairwise F_{st} values ((measure of pairwise population differentiation due to genetic structure)) are presented above the diagonal (single italicized value [region West Central] is not statistically significant). The total number of haplotypes from each region is given in parentheses in the row and column headers. The number of haplotypes shared among pairs of locations over the total number of haplotypes from a given pair of regions is presented below the diagonal

That is, all F_{st} values (measure of pairwise population differentiation due to genetic structure) were statistically significant except one, indicating restricted gene flow among the sampled regions (Table 2). The average pairwise genetic distances between ticks from the Hudson River Valley and ticks from any other of the sampled regions within New York were distinctly larger than all other between-region comparisons (Fig. 2a). Notably, ticks from the Hudson River Valley region were most distinct from ticks collected in the regions that were most distant to the Hudson River Valley (West and West Central regions).

The population genetic structure among ticks from the different regions of New York was driven by differences in the relative frequency of haplotypes rather than by differences in which haplotypes are present. Ticks from New York can be divided into six major phylogenetic clades which do not strictly reflect geography (Fig. 3a). At least one tick from each clade was collected from each geographic region except Clade 1, which was not represented in the West region (Fig. 3b). Similarly, regions shared high proportions of haplotypes (Table 2) and clusters of haplotypes (determined by K-means clustering) were represented across all regions for all values of K tested (Additional file 3: Figure S1). However, the relative frequency of ticks from each phylogenetic clade, K-means clusters and haplotypes varies among regions. For example, the Hudson River Valley had approximately equal proportions of ticks from each phylogenetic clade while the North region was dominated by ticks from two clades.

At a finer geographic scale, ticks collected from the same site were often more closely related to each other than to ticks from different sites (Fig. 3c). There was also some evidence of a correlation between geographic distance between sites where ticks were collected and genetic distance between those ticks (Mantel test, $P=0.061$), although the relationship is weak ($r=0.032$).

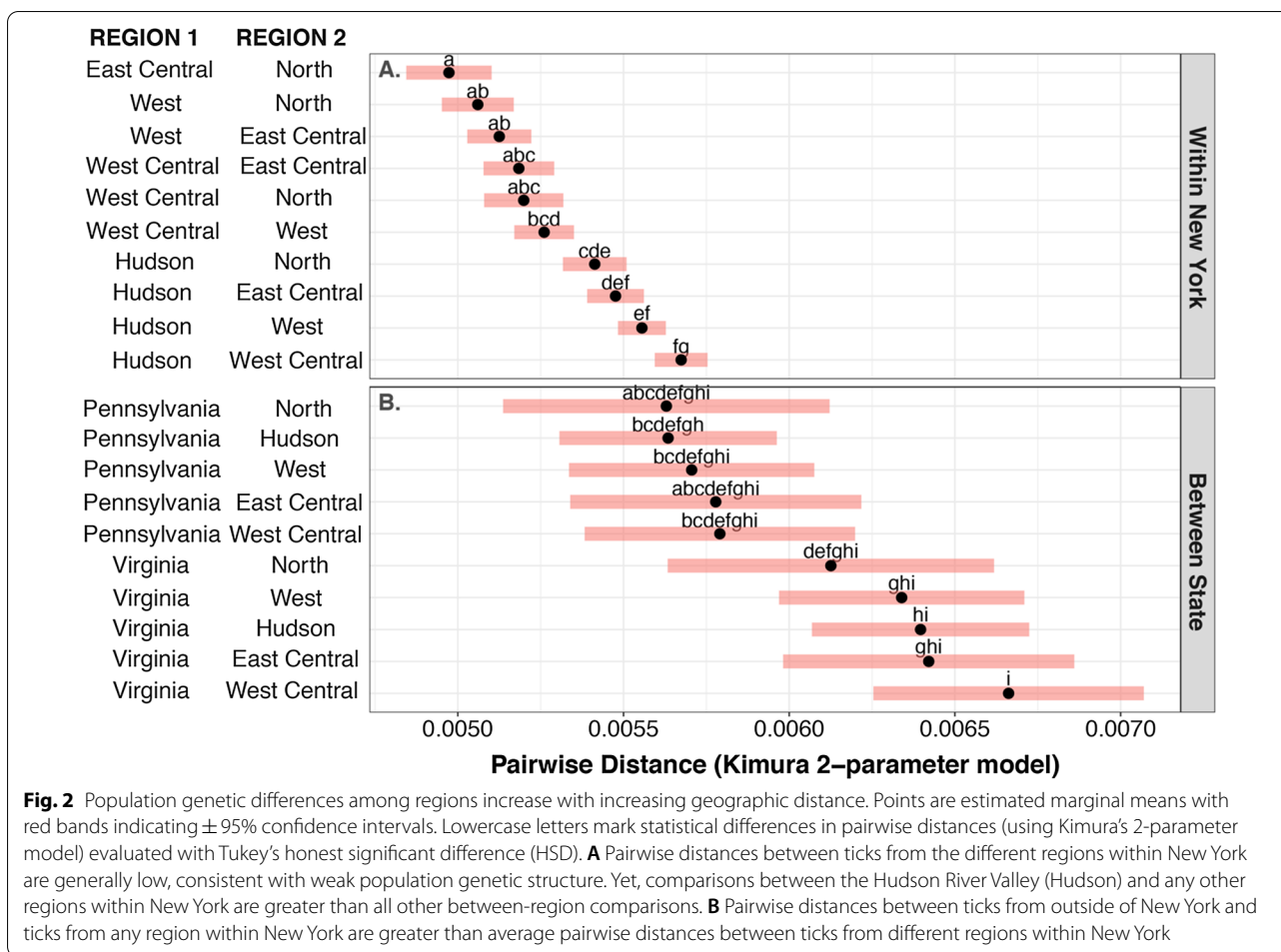


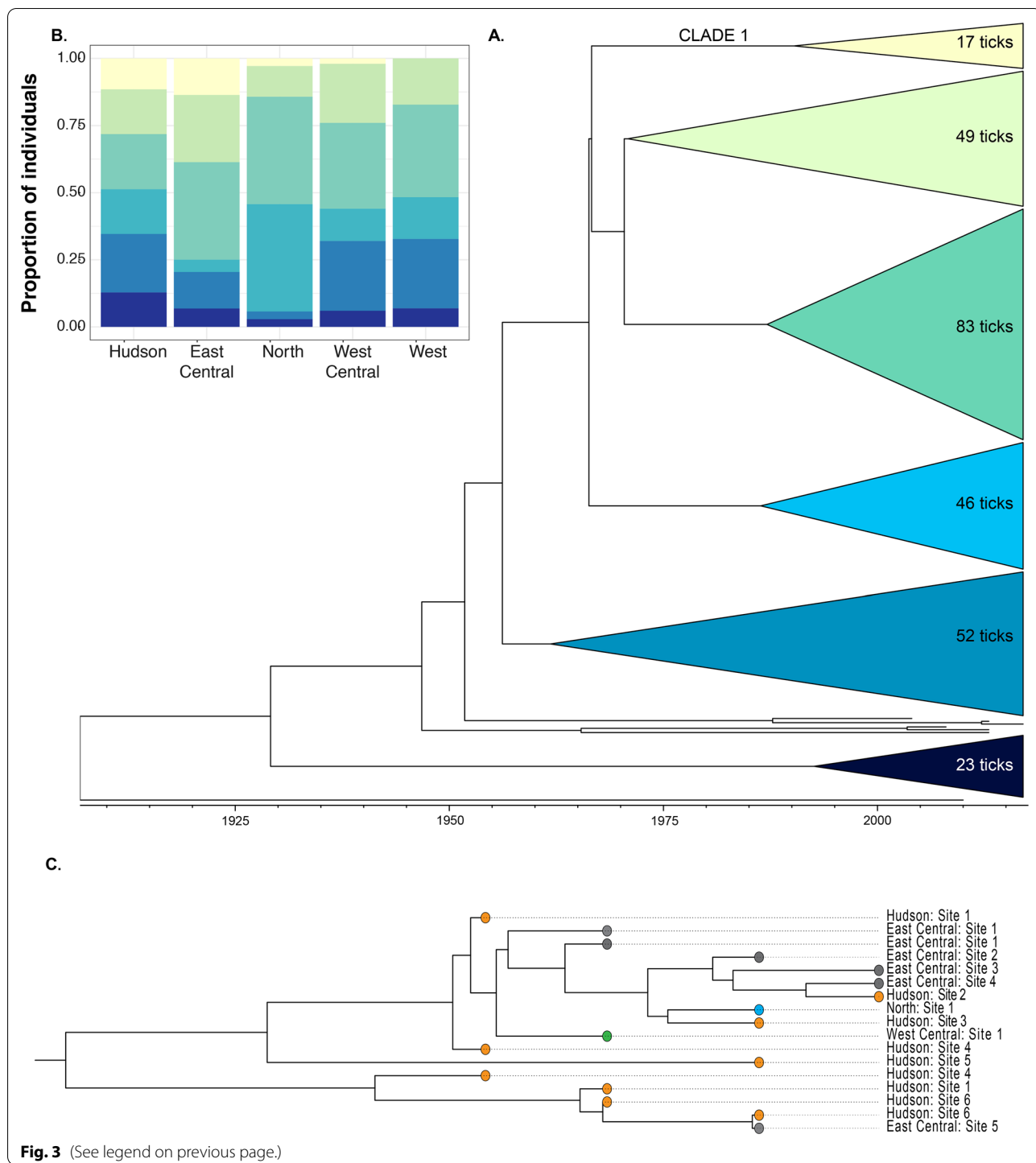
Fig. 2 Population genetic differences among regions increase with increasing geographic distance. Points are estimated marginal means with red bands indicating $\pm 95\%$ confidence intervals. Lowercase letters mark statistical differences in pairwise distances (using Kimura’s 2-parameter model) evaluated with Tukey’s honest significant difference (HSD). **A** Pairwise distances between ticks from the different regions within New York are generally low, consistent with weak population genetic structure. Yet, comparisons between the Hudson River Valley (Hudson) and any other regions within New York are greater than all other between-region comparisons. **B** Pairwise distances between ticks from outside of New York and ticks from any region within New York are greater than average pairwise distances between ticks from different regions within New York

Population genetic structure within and beyond New York
 Isolation by distance was further supported in analyses that included ticks collected in the states of Pennsylvania and Virginia. The average pairwise genetic distances between ticks from different regions was positively correlated with the geographic distances between those regions. Specifically, average pairwise genetic distances

were highest between ticks collected in regions of New York and those collected from Virginia; the second highest pairwise genetic distances were between ticks collected in regions of New York and those collected from Pennsylvania (Fig. 2b). When ticks from outside of New York were included in the analyses, the correlation between geographic distance and genetic distance had

(See figure on next page.)

Fig. 3 Relative abundance, not composition, of haplotypes of *I. scapularis* varies among regions. **A** Although closely related ticks could be found across New York, ticks within each clade tended to cluster geographically, suggesting isolation by distance at fine spatial scales. Dated maximum clade credibility (MCC) phylogeny of the best-fitting model inferred with BEAST (strict molecular clock model; Bayesian extended skyline population size) of *I. scapularis* in NY. Tips are collapsed into major clades in order to depict coarse scale patterns. Each clade is labeled with the number of ticks it contains. **B** Bar chart depicts the proportion of ticks from each clade found in each region. At least one tick from each clade was present within each region except that the Clade 1 is not present in the North region. Importantly, the proportion of ticks from each clade varied across New York regions. Ticks collected in the Hudson River Valley were distributed relatively evenly among the phylogenetic clades, whereas ticks from only two clades dominated in the North region. **C** Although closely related ticks could be found across New York, ticks within each clade tended to cluster geographically, suggesting isolation by distance at fine spatial scales. Clade 1 of the dated phylogeny of *I. scapularis* in New York with tips are colored by sampling region. In this example, ticks from four of the five regions were represented with unmistakable geographic clustering. That is, ticks from the East Central region were most closely related to each other while ticks from the Hudson River Valley were most closely related to each other



more statistical support (Mantel test $P=0.042$; $r=0.045$). While one tick from Pennsylvania was an outgroup to the phylogeny of all ticks included in this study, the other ticks from outside New York were not. These ticks represented unique haplotypes but shared SNPs with ticks

sampled from New York, consistent with a recent shared history of all sampled ticks.

Within-region genetic diversity varied across space and time. Such a pattern may suggest that the relative age of tick populations varies among regions. For example, the pairwise genetic distances between mitochondrial

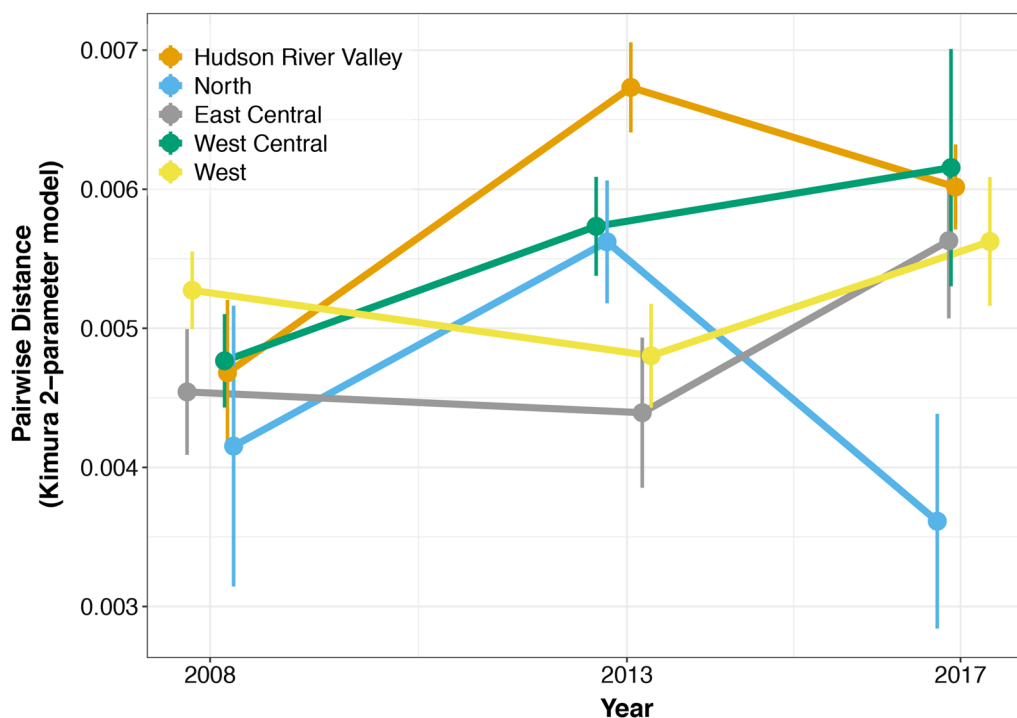


Fig. 4 Within-region population genetic diversity varies across space and time. Points represent the average pairwise distance (using the Kimura 2-parameter model) between ticks within each of the five regions of New York in each of three collection years. Vertical lines around each point represent $\pm 95\%$ confidence intervals and lines connect points from the same region across sampling years. The year 2004 was excluded because ticks were sampled only in the Hudson River Valley that year. Points are jittered on the X-axis for clarity. In 2008, the first year in which ticks were collected from all regions, the population genetic diversity did not differ among regions. Within-region diversity was generally high in the Hudson River Valley in subsequent years, and clearly higher than other regions among ticks collected in 2013. While it is clear that within-region diversity was lowest in the North region in the most recent sampling year, that pattern was inconsistent in previous years. Among all regions except for the North region, there was clear support for genetic diversity within regions increasing over time

sequences of ticks within the Hudson River Valley were generally higher than those of other regions, with the clearest support for this among ticks collected in 2013, suggesting tick populations from the Hudson River Valley may be the oldest in New York (Fig. 4). The greater diversity within the Hudson River Valley could also result from higher rates of migration into this region, among several other ecological or evolutionary factors. In contrast, the pairwise genetic distances between ticks in the North region were the lowest. While within-region diversity was lowest in the North region in the most recent sampling year (2017), that pattern was inconsistent in previous years. Among all regions except for the North region, diversity within regions appeared to be increasing over time.

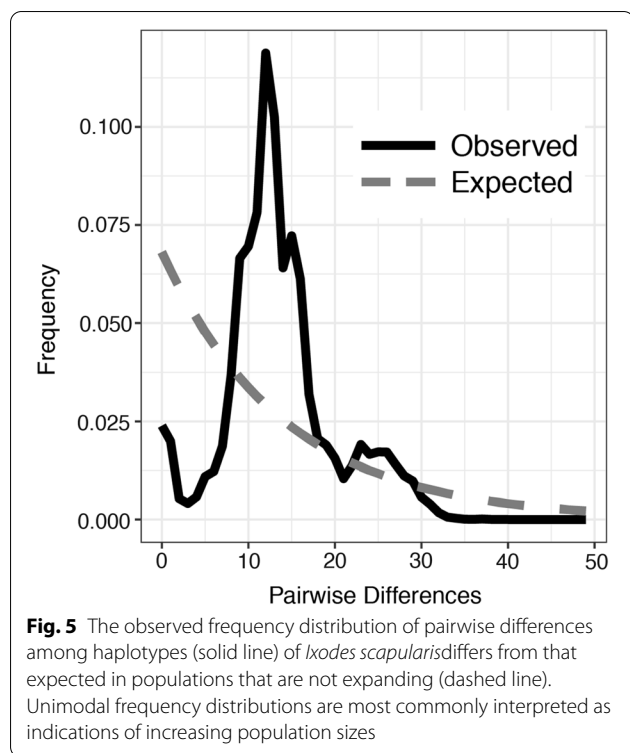
Population demography

The blacklegged tick population size has increased dramatically in the last two decades. The distribution of mismatches found among all sampled ticks (Fig. 5) and among ticks within each region differed from the

distribution that would be expected in populations experiencing no demographic changes. The unimodal distribution of pairwise differences is commonly interpreted as an increasing population size although they can also indicate increasing geographic range [62].

Discussion

We detailed the population dynamics of the blacklegged tick (*Ixodes scapularis*) in New York during a time frame in which tick-borne diseases were increasing in range and incidence. We found evidence of population genetic structure at a broad geographic scale due to differences in the relative abundance, but not the composition, of haplotypes among sampled ticks (Figs. 2a, 3b; Table 2). Said otherwise, genetically similar ticks were found across New York, suggesting that populations across the state are connected by migration and colonization events, but these events do not occur with sufficient regularity to homogenize the genetic diversity among populations. Limited gene flow among regions could be caused by rare migration events between regions or rare colonization



events due to ecological differences among regions. Limited gene flow resulted in a pattern of isolation by distance which was statistically supported when ticks from regions outside New York were included in analyses (Fig. 2b). Within most regions, the population genetic diversity increased over time, suggesting gene flow in the recent past. These data provide a foundation on which to generate testable hypotheses on the drivers of tick population dynamics occurring at finer scales.

Short- and long-range migration events are relatively common among ticks in New York. The active dispersal capabilities of ticks are minimal such that migration occurs primarily while attached to a vertebrate host [63–65]. Prior analyses of data from the Hudson River Valley (Hudson region) indicated that the probability that a tick population would colonize a novel location was correlated with how geographically close that location was to an already established population, consistent with gene flow via terrestrial vertebrate hosts [10]. The data presented here provide evidence of geographic signatures in phylogenetic relationships that support this conclusion; that is, ticks were often most closely related to ticks from the same and nearby collection sites. For example, ticks from the East Central region clustered together on the most probable phylogeny, and at a finer scale, ticks collected from Site 6 in the Hudson region are closely related to each other (Fig. 3c). Additionally, a Mantel test supports a correlation between genetic distance between

ticks and geographic distance between the sites where those ticks were collected.

Longer-range gene flow has also occurred, as some ticks were most closely related to ticks from distant sites, although this appears to be less common than more proximal gene flow (Fig. 3c). These findings are consistent with those of other studies in the continental USA which have demonstrated that migrating birds impact tick range expansion and the maintenance of connectivity among tick populations at broad spatial scales [66, 67]. While an alternative cause of closely-related ticks from geographically distant areas could be multiple short-distance gene flow events by ticks attached to terrestrial reservoir hosts, the time frame represented by these data is too short to support this premise.

Analyses of this coarse scale dataset do not indicate the ecological or environmental processes that have driven tick range expansion across New York. Investigations of ticks migrating into Canada suggest that songbirds carry ticks northward beyond areas with climates suitable for tick populations [68, 69]. In such areas, tick populations cannot persist, and those areas are repeatedly colonized by migrating birds carrying ticks [68–70]. While birds are likely involved in dispersing ticks among locations within New York, there is no evidence that migration of tick-infested birds is driving tick range expansions. Differences in the drivers of ranges expansion within New York, which is not at the current northern boundary of the species range, and into Canada, which is at the species range boundary, should be expected as the processes impacting tick population dynamics are likely to vary among regions with different underlying ecologies and environmental conditions [71].

Survey-based surveillance efforts have revealed which landscape and climatic features drive tick presence and absence, tick densities and subsequent tick-borne disease risk [22, 72]. Mitochondrial genome sequencing allows for the study of gene flow between established populations, which is not possible with epidemiological survey data, to determine how environmental features impact current migration as well as range expansions. In the present study, mitochondrial genome sequencing revealed the role of short- and long-range migration events in the evolutionary history of ticks in New York. Analyses of entire mitochondrial genomes increase analytical power over many prior studies that use a few loci with limited genetic variation [10, 73], thus permitting investigations at fine spatiotemporal scales, the scale at which many ecological processes occur. For example, hypotheses that focus on identifying and quantifying the impact of environmental features on migration rates, colonization rates, population persistence and on other tick population dynamic characteristics can be tested.

Conclusions

Understanding tick population dynamics is critical to the design of effective strategies to mitigate risk from tick-borne pathogens. Tick-borne pathogens require the establishment of populations of highly competent vectors, like *I. scapularis*, to occur at epidemiologically relevant levels in a given area [69, 74]. While there is evidence of some limitations to gene flow, tick migration occurs at both coarse and fine scales, which could have implications for further range expansion as habitat suitability for ticks changes due to global climate change [75, 76]. Analyses of mitochondrial genome sequencing data will expound upon previously identified drivers of tick presence and abundance and can further identify additional drivers. These analyses are critical to ascertain the connectivity among tick populations, how and why current gene flow occurs between established populations, and how environmental features likely impact these population dynamics.

Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s13071-022-05304-9>.

Additional file 1: Table S1. Sample metadata. For each tick, this includes year collected, sex, and collection site information.

Additional file 2: Table S2. Long-range PCR primer sequences.

Additional file 3: Figure S1. Clusters of haplotypes (determined by K-means clustering) were represented across all regions for all values of K tested. Each row represent a different value for K (from K=6 to K=9). For each value of K, the proportion of each cluster (colored along a gray gradient) within each region is shown.

Author contributions

KO, ZO, and DB organized and designed the study. MP, RF, JO, JH, LS, and PB organized and performed the fieldwork and subsequent DNA extractions. ZO designed primers for sequencing mitochondrial genome, and ZO and KO performed the amplification and sequencing of mitochondrial genomes. KO analysed the data and drafted the manuscript. All authors read and approved the final manuscript.

Funding

National Institutes of Health, R01AI142572, R01AI142572, R01AI142572, R01AI142572, R01AI142572, R01AI142572, R01AI142572, R01AI142572, R01AI142572, R01AI142572, Burroughs Wellcome Fund, 1012376, 1012376, Centers for Disease Control and Prevention, U01CK000509, U01CK000509, U01CK000509, U01CK000509, U01CK000509, U01CK000509

Availability of data and materials

Sequence data that support the findings of this study were deposited in GenBank (Accession PRJNA825593).

Declarations

Competing interests

The authors declare no competing interests.

Author details

¹Department of Biology, University of Pennsylvania, Philadelphia, PA, USA. ²New York State Department of Health, Albany, NY, USA. ³Department of Health, Central New York Regional Office, Syracuse, NY 13202, USA. ⁴Paul Smith’s College, Paul Smiths, NY, USA.

Received: 4 March 2022 Accepted: 15 April 2022

Published online: 28 June 2022

References

- Jones KE, Patel NG, Levy MA, Storeygard A, Balk D, Gittleman JL, et al. Global trends in emerging infectious diseases. *Nature*. 2008;451:990–3.
- Taylor LH, Latham SM, Woolhouse ME. Risk factors for human disease emergence. *Philos Trans R Soc Lond B Biol Sci*. 2001;356:983–9.
- Patz JA, Olson SH, Uejio CK, Gibbs HK. Disease emergence from global climate and land use change. *Med Clin North Am*. 2008;92:1473–91.
- Steiger DBM, Ritchie SA, Laurance SGW. Mosquito communities and disease risk influenced by land use change and seasonality in the Australian tropics. *Parasit Vectors*. 2016;9:387.
- Baylis M. Potential impact of climate change on emerging vector-borne and other infections in the UK. *Environ Health*. 2017;16:112.
- Liang L, Gong P. Climate change and human infectious diseases: a synthesis of research findings from global and spatio-temporal perspectives. *Environ Int*. 2017;103:99–108.
- Baeza A, Santos-Vega M, Dobson AP, Pascual M. The rise and fall of malaria under land-use change in frontier regions. *Nat Ecol Evol*. 2017;1:108.
- Samy AM, Elaagip AH, Kenawy MA, Ayres CFJ, Peterson AT, Soliman DE. Climate change influences on the global potential distribution of the mosquito *Culex quinquefasciatus*, vector of West Nile virus and lymphatic filariasis. *PLoS ONE*. 2016;11:e0163863.
- Ogden NH, Lindsay LR. Effects of climate and climate change on vectors and vector-borne diseases: ticks are different. *Trends Parasitol*. 2016;32:646–56.
- Khatchikian CE, Prusinski MA, Stone M, Backenson PB, Wang I-N, Foley E, et al. Recent and rapid population growth and range expansion of the Lyme disease tick vector, *Ixodes scapularis*, in North America. *Evolution*. 2015;69:1678–89.
- Vittor AY, Pan W, Gilman RH, Tielsch J, Glass G, Shields T, et al. Linking deforestation to malaria in the Amazon: characterization of the breeding habitat of the principal malaria vector, *Anopheles darlingi*. *Am J Trop Med Hyg*. 2009;81:5–12.
- Khatchikian CE, Prusinski M, Stone M, Bryon Backenson P, Wang IN, Levy MZ, et al. Geographical and environmental factors driving the increase in the Lyme disease vector *Ixodes scapularis*. *Ecosphere*. 2012;3:1–18.
- Humphrey PT, Caporale DA, Brisson D. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution*. 2010;64:2653–63.
- Kyle JL, Harris E. Global spread and persistence of dengue. *Annu Rev Microbiol*. 2008;62:71–92.
- Martens WJ, Niessen LW, Rotmans J, Jetten TH, McMichael AJ. Potential impact of global climate change on malaria risk. *Environ Health Perspect*. 1995;103:458–64.
- Hamer SA, Tsao JI, Walker ED, Hickling GJ. Invasion of the Lyme disease vector *Ixodes scapularis*: implications for *Borrelia burgdorferi* endemicity. *EcoHealth*. 2010;7:47–63.
- Kannagara DW, Sidra S, Pritiben P. First case report of inducible heart block in Lyme disease and an update of Lyme carditis. *BMC Infect Dis*. 2019;19:428.
- Humphrey PT, Caporale DA, Brisson D. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution*. 2010;64:2653–63.
- Van Zee J, Piesman JF, Hojgaard A, Black WC IV. Nuclear markers reveal predominantly north to south gene flow in *Ixodes scapularis*, the tick vector of the Lyme disease spirochete. *PLoS ONE*. 2015;10:e0139630–e0139630.
- Eisen RJ, Eisen L, Beard CB. County-Scale Distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the Continental United States. *J Med Entomol*. 2016;53:349–86.
- Sonenshine DE. Range expansion of tick disease vectors in North America: implications for spread of tick-borne disease. *Int J Environ Res Public Health*. 2018;15:478.
- Tran T, Prusinski MA, White JL, Falco RC, Vinci V, Gall WK, et al. Spatio-temporal variation in environmental features predicts the distribution and abundance of *Ixodes scapularis*. *Int J Parasitol*. 2021;51:311–20.

23. Schaub EM, Gertz SJ, Maple WT, Ostfeld RS. Coinfection of blacklegged ticks (Acari: Ixodidae) in Dutchess County, New York, with the agents of Lyme disease and human granulocytic ehrlichiosis. *J Med Entomol.* 1998;35:901–3.
24. Dolan MC, Maupin GO, Schneider BS, Denatale C, Hamon N, Cole C, et al. Control of immature *Ixodes scapularis* (Acari: Ixodidae) on rodent reservoirs of *Borrelia burgdorferi* in a residential community of Southeastern Connecticut. *J Med Entomol.* 2004;41:1043–54.
25. Pepin KM, Eisen RJ, Mead PS, Piesman J, Fish D, Hoen AG, et al. Geographic variation in the relationship between human Lyme disease incidence and density of infected host-seeking *Ixodes scapularis* nymphs in the Eastern United States. *Am J Trop Med Hyg.* 2012;86:1062–71.
26. Perret J-L, Rais O, Gern L. Influence of climate on the proportion of *Ixodes ricinus* nymphs and adults questing in a tick population. *J Med Entomol.* 2004;41:361–5.
27. Oliver J, Means RG, Kogut S, Prusinski M, Howard JJ, Layne LJ, et al. Prevalence of *Borrelia burgdorferi* in small mammals in New York State. *J Med Entomol.* 2006;43:924–35.
28. Barbour AG, Fish D. The biological and social phenomenon of Lyme disease. *Science.* 1993;260:1610–6.
29. Brisson D, Dykhuizen DE, Ostfeld RS. Conspicuous impacts of inconspicuous hosts on the Lyme disease epidemic. *Proc Biol Sci.* 2008;275:227–35.
30. Falco RC, McKenna DF, Daniels TJ, Nadelman RB, Nowakowski J, Fish D, et al. Temporal relation between *Ixodes scapularis* abundance and risk for Lyme disease associated with erythema migrans. *Am J Epidemiol.* 1999;149:771–6.
31. Huang H, Fang M, Jostins L, Mirkov MU, Boucher G, Anderson CA, et al. Fine-mapping inflammatory bowel disease loci to single-variant resolution. *Nature.* 2017;547:173–8.
32. Ulirsch JC, Nandakumar SK, Wang L, Giani FC, Zhang X, Rogov P, et al. Systematic functional dissection of common genetic variation affecting red blood cell traits. *Cell.* 2016;165:1530–45.
33. Ulirsch JC, Lareau CA, Bao EL, Ludwig LS, Guo MH, Benner C, et al. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nat Genet.* 2019;51:683–93.
34. Betts MG, Hadley AS, Frey DW, Frey SJK, Gannon D, Harris SH, et al. When are hypotheses useful in ecology and evolution? *Ecol Evol.* 2021;11:5762–76.
35. Gandomi A, Haider M. Beyond the hype: big data concepts, methods, and analytics. *Int J Inf Manage.* 2015;35:137–44.
36. Ginsberg HS, Ewing CP. Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, *Ixodes dammini*, and lone-star ticks, *Amblyomma americanum* (Acari: Ixodidae). *Exp Appl Acarol.* 1989;7:313–22.
37. Prusinski MA, Kokas JE, Hukey KT, Kogut SJ, Lee J, Backenson PB. Prevalence of *Borrelia burgdorferi* (Spirochaetales: Spirochaetaceae), *Anaplasma phagocytophilum* (Rickettsiales: Anaplasmataceae), and *Babesia microti* (Piroplasmida: Babesiidae) in *Ixodes scapularis* (Acari: Ixodidae) collected from recreational lands in the Hudson Valley Region, New York State. *J Med Entomol.* 2014;51:226–36.
38. Durden L, Keirans J. Nymphs of the genus *Ixodes* (Acari: Ixodidae) of the United States: taxonomy, identification key, distribution, hosts, and medical/veterinary importance. Thomas Say Publications in Entomology: Monographs. Lanham: Entomological Society of America; 1996.
39. Keirans JE, Clifford CM. The genus *Ixodes* in the United States: a scanning electron microscope study and key to the adults. *J Med Entomol.* 1978;15:1–38.
40. Devevey G, Brisson D. The effect of spatial heterogeneity on the aggregation of ticks on white-footed mice. *Parasitology.* 2012;139:915–25.
41. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics.* 2012;13:134.
42. Martin M. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J.* 2011;17:10–2.
43. Montagna M, Sasser D, Griggio F, Epis S, Bandi C, Gissi C. Tick-Box for 3'-end formation of mitochondrial transcripts in *Ixodida* Basal Chelicerates and *Drosophila*. *PLoS ONE.* 2012;7:e47538.
44. Price DC, Brennan JR, Wagner NE, Egizi AM. Comparative hologenomics of two *Ixodes scapularis* tick populations in New Jersey. *PeerJ.* 2021;9:e12313.
45. Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. *Bioinforma.* 2010;26:589–95.
46. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv: Genomics. 2013.
47. Broad Institute, GitHub Repository. Picard Toolkit. 2019. <http://broadinstitute.github.io/picard/>. Accessed 22 Feb 2022.
48. Poplin R, Ruano-Rubio V, DePristo MA, Fennell TJ, Carneiro MO, der Auwera GA, et al. Scaling accurate genetic variant discovery to tens of thousands of samples. *bioRxiv.* 2018;10:1004450.
49. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytzky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.* 2010;20:1297–303.
50. Bouckaert R, Vaughan TG, Barido-Sottani J, Duchêne S, Fourment M, Gavryushkina A, et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLOS Comput Biol.* 2019;15:e1006650.
51. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using tracer 1.7. *Syst Biol.* 2018;67:901–4.
52. Lartillot N, Philippe H. Computing Bayes factors using thermodynamic integration. *Syst Biol.* 2006;55:195–207.
53. Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol.* 2007;7:214.
54. Librado P, Rozas J. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics.* 2009;25:1451–2.
55. Wickham H. ggplot2: Elegant graphics for data analysis. New York: Springer-Verlag; 2016.
56. Kimura M. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. *J Mol Evol Springer.* 1980;16:111–20.
57. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol Biol Evol.* 2018;35:1547.
58. Lenth R. Emmeans: Estimated marginal means, aka least-squares means. R Package version 1.1.3. 2018. <https://CRAN.R-project.org/package=emmeans>. Accessed 22 Feb 2022.
59. Wright S. Isolation by distance. *Genetics.* 1943;28:114–38.
60. Mantel N. The detection of disease clustering and a generalized regression approach. *Cancer Res.* 1967;27:209–20.
61. Dixon P. VEGAN, a package of R functions for community ecology. *J Veg Sci.* 2003;14:927–30.
62. Rogers AR, Harpending H. Population growth makes waves in the distribution of pairwise genetic differences. *Mol Biol Evol.* 1992;9:552–69.
63. Falco RC, Fish D. Horizontal movement of adult *Ixodes dammini* (Acari: Ixodidae) attracted to CO₂-baited traps. *J Med Entomol.* 1991;28:726–9.
64. Leighton PA, Koffi JK, Pelcat Y, Lindsay LR, Ogden NH. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. *J Appl Ecol.* 2012;49:457–64.
65. Madhav NK, Brownstein JS, Tsao JI, Fish D. A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). *J Med Entomol.* 2004;41:842–52.
66. Cumbie AN, Heller EL, Bement ZJ, Phan A, Walters EL, Hynes WL, et al. Passerine birds as hosts for *Ixodes* ticks infected with *Borrelia burgdorferi* sensu stricto in southeastern Virginia. *Ticks Tick Borne Dis.* 2021;12:101650.
67. Tonelli BA, Dearborn DC. An individual-based model for the dispersal of *Ixodes scapularis* by ovenbirds and wood thrushes during fall migration. *Ticks Tick Borne Dis.* 2019;10:1096–104.
68. Ogden NH, Lindsay LR, Hanincová K, Barker IK, Bigras-Poulin M, Charron DF, et al. Role of migratory birds in introduction and range expansion of *Ixodes scapularis* ticks and of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. *Appl Environ Microbiol.* 2008;74:1780–90.
69. Ogden NH, Lindsay LR, Leighton PA. Predicting the rate of invasion of the agent of Lyme disease *Borrelia burgdorferi*. *J Appl Ecol.* 2013;50:510–8.
70. Ogden NH, Feil EJ, Leighton PA, Lindsay LR, Margos G, Mechai S, et al. Evolutionary aspects of emerging Lyme disease in Canada. *Appl Environ Microbiol.* 2015;81:7350–9.
71. Dong Y, Huang Z, Zhang Y, Wang YXG, La Y. Comparing the climatic and landscape risk factors for Lyme disease cases in the Upper Midwest and Northeast United States. *Int J Environ Res Public Health.* 2020;17:1548.
72. Talbot B, Slatculescu A, Thickstun CR, Koffi JK, Leighton PA, McKay R, et al. Landscape determinants of density of blacklegged ticks, vectors of Lyme

disease, at the northern edge of their distribution in Canada. *Sci Rep.* 2019;9:16652.

73. Araya-Anchetta A, Busch JD, Scoles GA, Wagner DM. Thirty years of tick population genetics: a comprehensive review. *Infect Genet Evol.* 2015;29:164–79.
74. Oliver JD, Bennett SW, Beati L, Bartholomay LC. Range Expansion and Increasing *Borrelia burgdorferi* Infection of the Tick *Ixodes scapularis* (Acari: Ixodidae) in Iowa, 1990–2013. *J Med Entomol.* 2017;54:1727–34.
75. Wang J, Guan Y, Wu L, Guan X, Cai W, Huang J, et al. Changing lengths of the four seasons by global warming. *Geophys Res Lett.* 2021;48:1–9.
76. Intergovernmental Panel on Climate Change (IPCC). Climate change 2021: the physical science basis. Contribution of Working Group I to the sixth assessment report of the Intergovernmental Panel on Climate Change. 2021. <https://www.ipcc.ch/report/ar6/wg1/>. Accessed 22 Feb 2022.

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

