doi:10.1111/evo.12690

Recent and rapid population growth and range expansion of the Lyme disease tick vector, *Ixodes scapularis,* in North America

Camilo E. Khatchikian,[1,2] Melissa A. Prusinski,[3] Melissa Stone,[4] Peter Bryon Backenson,[3,4] Ing-Nang Wang,[4] Erica Foley,[1] Stephanie N. Seifert,[1] Michael Z. Levy,[1] and Dustin Brisson[1,5]

[1]*University of Pennsylvania, Philadelphia, Pennsylvania 19104*

[2]*E-mail: ckhatchikian@gmail.com*

[3]*New York Department of Health, Albany, New York 12237*

[4]*State University of New York, Albany, New York 12222*

[5]*E-mail: dbrisson@upenn.edu*

Received April 30, 2014
Accepted March 31, 2015

Migration is a primary force of biological evolution that alters allele frequencies and introduces novel genetic variants into populations. Recent migration has been proposed as the cause of the emergence of many infectious diseases, including those carried by blacklegged ticks in North America. Populations of blacklegged ticks have established and flourished in areas of North America previously thought to be devoid of this species. The recent discovery of these populations of blacklegged ticks may have resulted from either in situ growth of long-established populations that were maintained at very low densities or by migration and colonization from established populations. These alternative evolutionary hypotheses were investigated using Bayesian phylogeographic approaches to infer the origin and migratory history of recently detected blacklegged tick populations in the Northeastern United States. The data and results indicate that newly detected tick populations are not the product of in situ population growth from a previously established population but from recent colonization resulting in a geographic range expansion. This expansion in the geographic range proceeded primarily through progressive and local migration events from southern populations to proximate northern locations although long-distance migration events were also detected.

KEY WORDS: **Blacklegged tick, emerging disease, phylogeography, zoonosis.**

Migration across geographical space is a major evolutionary force that alters allele frequencies within population. The rate, direction, and timing of migration can maintain genetic variation over time and space as well as expose genetic variants to novel selective environments. Recent migration of arthropod vectors of infectious diseases has been implicated in the emergence of many emerging and re-emerging diseases (Lounibos 2002; Kaplan et al. 2010; Jaenson and Lindgren 2011; Foley et al. 2013; Khatchikian et al. 2015). For example, the increases in the geographic range and population sizes of several tick species over the previous few decades in the northern hemisphere have resulted in dramatic increases in the incidence of tick-borne diseases such as tick-borne encephalitis, Lyme disease, babesiosis, anaplasmosis, and human granulocytic Ehrlichiosis. Although currently prevalent throughout much of North America and Europe, neither the direction, rate, and timing of migration, nor the demographic histories of many emerging tick populations have been investigated on a temporal and spatial scale relevant to the recent population dynamics. In this study, we assess the demographic dynamics and the direction, rates, and timing of migration of *Ixodes scapularis* tick populations that have recently expanded in size or geographic range in the Northeastern United States.

Data Archival Location: DNA sequences are deposited in GenBank and accession numbers reported in tables and the data archiving section of the manuscript.

EVOLUTION
INTERNATIONAL JOURNAL OF ORGANIC EVOLUTION

The blacklegged tick, *I. scapularis*, is the most important vector of human diseases in Northeastern and Midwestern United States (Hamer et al. 2010). Further, increases in the geographic range in the human incidence of diseases vectored by *I. scapularis* over the last decades parallel the apparent increases in the geographic range of tick populations (Ogden et al. 2009; Brisson et al. 2010; Hamer et al. 2010; Khatchikian et al. 2012; Leighton et al. 2012; Kelly et al. 2014). It is not clear, however, if the recently detected blacklegged tick populations in areas once thought to be beyond the geographic range of *I. scapularis* are the result of recent increases in the density of established populations from below detectable densities or are the consequence of immigration into new areas.

The importance of population growth and progressive and long-distance migration to the apparent range expansion of *I. scapularis* can be established using serial coalescent and phylogeographic modeling approaches. These approaches allow for the identification of migratory events and the reconstruction of the demographic history of populations to reveal the colonization processes. Further, the timing of historical colonization events can be estimated using heterochronous samples (samples collected at multiple time points) in a coalescent framework (Firth et al. 2010). The accuracy of these statistical approaches in assessing the demographic history far exceeds direct field estimates when these are of an erratic temporal and spatial nature, as is the case for most collections. Although the demographic and migratory history of *I. scapularis* populations in the United States has been estimated on geological time scales (Humphrey et al. 2010), the recent demographic dynamics and the rate, direction, and timing of migration that can cause evolutionary changes in these populations have not been explored.

In the current study, we analyze fine-scale, temporally structured, and spatially structured *I. scapularis* samples collected in the Hudson River Valley, New York, during the recent increase in population densities or geographic range. *Ixodes scapularis* populations have been endemic to the southern areas of the Hudson River Valley for several decades whereas populations in the northern areas of the sampling region have only been detected in the last decade (Daniels et al. 1998; Khatchikian et al. 2012). Genetic sequence data from these samples were analyzed in a coalescent framework to test three specific hypotheses. First, is the geographic range of ticks expanding through migration or are have populations been detected only recently due to recent in situ growth from small endemic populations? Second, does the rate of gene flow scale with geographic distance? Lastly, is gene flow unidirectional (from established large populations to smaller populations) or multidirectional? The hypothesis testing framework of this study provides a robust characterization of the evolutionary processes driving the apparent range expansion of *I. scapularis* ticks in North America.

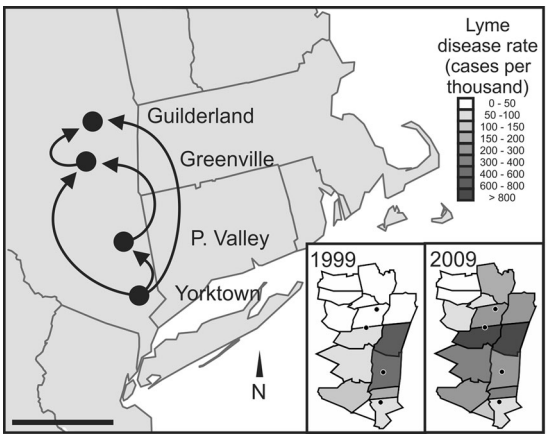

Figure 1. **Illustration of the south-to-north range expansion inferred in this study using samples collection at four sites (named after their closest township). Arrows indicate migratory patterns with strong Bayesian posterior support. Scale bar represents 100 km. Insets represent countywide Lyme disease incidence rates in 1999 and 2009 (Khatchikian et al. 2012). The collection sites are indicated with black dots.**

Table 1. **Genetic diversity indices including the number of haplotypes (*H*) and haplotype diversity (H_d,) for each sampled location.**

		Sampled location			
		Guilderland	Greenville	Pleasant Valley	Yorktown
Sample size		0/20/20	20/20/20	20/20/20	20/20/20
16S	*H*	7	9	9	10
	H_d	0.58	0.67	0.72	0.79
COII	*H*	13	18	22	15
	H_d	0.80	0.89	0.89	0.87
CR	*H*	12	24	23	16
	H_d	0.81	0.93	0.93	0.89
Concatenated	*H*	21	41	43	35
	H_d	0.86	0.98	0.98	0.96

Sample sizes for each collection year at each site are presented as 2004/2007/2009.

Methods

POPULATION SAMPLING AND SAMPLE PREPARATION

Ixodes scapularis ticks were collected in four locations (hereafter named after the closest township, Fig. 1; Table 1) in the Hudson River Valley, New York (USA) in 2004, 2007, and 2009. In the northernmost location, Guilderland, no ticks were collected in 2004. For each collection year, subsets of 20 ticks were randomly selected from each location resulting in 220 individual ticks processed. Host-seeking ticks were collected using standardized

flagging methods (Ginsberg and Ewing 1989) as part of routine tick-borne pathogen surveillance by the New York State Department of Health. Collected ticks were placed into individual 1.5 mL Eppendorf tubes and preserved in 80% EtOH and stored at −20°C until nucleic acid extraction. The majority of the individuals were adults (82% vs. 18% nymphs).

Genomic DNA was extracted using standard genomic DNA purification protocols. Briefly, individual ticks were double rinsed with nuclease-free ultra-pure water and placed in individual 2.0 mL round-bottomed Safe-Lock™ tubes (Eppendorf North America, Inc., Hauppauge, NY) with two 4 mm stainless steel grinding beads and 180 μl PBS, pH 7.2. Samples were pulverized using an electric tissue homogenizer (MM 300, Retsch, Haan, Germany). Whole genomic DNA was extracted using the DNeasy™ Blood and Tissue kit (Qiagen, Valencia, CA) following the manufacturer's recommendations (Purification of Total DNA from Insects). DNA extraction was conducted in a class 2 biological safety cabinet with laminar airflow (SterilGARD™III Advance, Baker Co., Sanford, ME) designated exclusively for DNA extraction, and sterile aerosol-barrier tips were used during all procedures to prevent cross-contamination. DNA from individual ticks was eluted in 200 μl ultra-pure water and stored at −20°C until further analysis.

DNA SEQUENCING

Three mitochondrial loci—the 16S ribosomal gene, cytochrome oxidase subunit II gene (COII), and the mitochondrial control region (CR)—were amplified by polymerase chain reaction (PCR). Primer sequences and cycling conditions are reported in Appendix S1 and Table S1 (Simon et al. 1994; Casati et al. 2008). PCR contamination was avoided by preparing reactions in a biological safety cabinet, using barrier tips, and using at least one negative control for every eight reactions. Amplicons were purified using PCR cleaning columns and sequenced using dye terminator cycle sequencer technology at the sequencing facilities of the University of Pennsylvania. The amplification and sequencing success rate was 83%. All unsuccessfully amplified or sequenced fragments, as well as all singleton mutations, were re-amplified and re-sequenced. Sequences were aligned using Geneious version 4.5 (Kearse et al. 2012). Any indels (insertions/deletions) detected in the alignments were removed due to ambiguities in their interpretation. The lengths of the final alignments were 686 bp, 507 pb, and 380 bp for the 16S, COII, and CR loci, respectively.

POPULATION GENETIC ANALYSES

The number of haplotypes (H), haplotype diversity (H_d) indices, and mismatch distributions was calculated using DNASP version 5 (Librado and Rozas 2009). Potential biases in population diversity estimators caused by the inequality in sample sizes between the northernmost location (40) and all other locations (60) was corrected for by using a rarefaction approach (H': algorithm provided by S. Holland and available at http://strata.uga.edu/software/) and using multiple random reduction to standardize haplotype diversity (H_d') to the smaller sample size using 10,000 subsamples (Leberg 2002). Evidence of population genetic structure was assessed using the F_{st} metric based on haplotype frequencies in Arlequin version 3.5.1.2 (Excoffier and Lischer 2010).

PHYLOGEOGRAPHIC INFERENCE

The phylogeographic history of the *I. scapularis* populations was inferred in Beast version 1.7.4 using a discrete spatial model (Drummond et al. 2012). Beast estimates the phylogenetic topology while simultaneously inferring the geographic location of ancestral nodes, migratory events among locations, and demography. We used the Bayesian stochastic search variable selection (BSSVS) approach (Lemey et al. 2009b) with conditional reference priors (Ferreira and Suchard 2008) to estimate migration among locations (the parameters in the nonsymmetrical connectivity matrix). The support for each parameter retained in the BSSVS was evaluated with Bayes factors using the ad hoc tool included in the Beast package (Suchard et al. 2005).

Estimates of migration events and the ancestral locations of haplotypes are informed by the demographic model parameters in the Beast algorithm. We compared the fit of the data to six demographic models including constant population size (Kingman 1982), expansion-growth (Griffiths and Tavare 1994), logistic growth (Griffiths and Tavare 1994), extended Bayesian skyline plot (linear and stepwise models; Heled and Drummond 2008), and Bayesian skyride (Minin et al. 2008). For each demographic model, the Bayesian reconstruction of the phylogeographic history of *I. scapularis* was estimated in Beast with a strict molecular clock using a Markov chain Monte Carlo algorithm run for 500,000,000 generations, sampled every 20,000 generations. The first 100 million generations were discarded as burn-in. The maximum clade credibility tree was calculated while keeping target heights using the TreeAnnotator included in the Beast package. The convergence of models and the acceptable effective sampling sizes for parameters were assessed in Tracer version 1.6 (Rambaut et al. 2014).

The recently developed path sampling (PS; Lartillot and Philippe 2006) and stepping-stone (SS; Xie et al. 2011) sampling (Xie et al. 2011) algorithms were used to estimate the marginal likelihoods of the data under each demographic model. The PS and SS algorithms outperform previous methods (Baele et al. 2012). Estimates of the marginal likelihoods by both algorithms for each model were determined using estimates derived from 2,000,000 samples of 100 path steps. The marginal likelihoods were used to evaluate the performance of each demographic model to estimate the phylogeographic history of *I. scapularis* in this region. In the best performing model, the minimum number of migration events

Table 2. Frequency of 16S haplotypes at each sampling location from each year of collection (2004/2007/2009) and GenBank accession numbers.

Haplotype number	Total	Yorktown	Pleasant Valley	Greenville	Guilderland	Accession numbers
1	88	7/8/4	6/8/5	2/11/12	0/12/13	KR092226
2	73	6/8/4	6/12/7	15/8/1	0/4/2	KR092223
3	17	1/0/3	2/0/3	0/0/3	0/2/3	KR092231
4	10	2/1/5	1/0/0	0/0/1	0/0/0	KR092230
5	8	3/0/1	1/0/2	0/0/0	0/1/0	KR092228
6	6	0/0/1	0/0/1	0/0/3	0/0/1	KR092235
7	4	0/0/0	1/0/2	1/0/0	0/0/0	KR092225
8	4	0/1/0	1/0/0	0/1/0	0/1/0	KR092232
9	3	0/1/1	0/0/0	0/0/0	0/0/1	KR092234
10	2	0/0/1	0/0/0	1/0/0	0/0/0	KR092224
11	2	1/1/0	0/0/0	0/0/0	0/0/0	KR092229
12	2	0/0/0	2/0/0	0/0/0	0/0/0	KR092233
13	1	0/0/0	0/0/0	1/0/0	0/0/0	KR092226

among branches, based on location inferences with high posterior support (>75%), were recorded along with the starting location and destination. For all nodes with high posterior support (>75%), the estimated branching time and 95% high-density probability (HDP) range were recorded. All identified migratory events were consistent with migratory events inferred using the BSSVS procedure. The robustness of the results was assessed by using a random tip operator to confirm the presence of the phylogenetical signal (Appendix S5).

POPULATION DEMOGRAPHY

The demographic parameters θ and the exponential growth parameter *g* were estimated with LAMARC version 2.1.10 (Kuhner 2006) using the Bayesian formulation. As Lamarc does not accommodate heterochronous datasets (see LAMARC documentation), we split our dataset by collection year; the 2004 dataset included 60 individuals from three locations and the 2007 and 2009 datasets each included 80 individuals from four locations. The mean values were calculated of the parameters estimated for each location. After initial exploratory runs, the prior for the growth parameter (*g*) was adjusted to a uniform distribution between −500 and 2000 and the prior for the migration parameter was adjusted to a uniform distribution between 0.01 and 10,000. Final analyses were run with two replicates with two chains automatically adjusting the heat for 10 million steps following a 2 million step burn-in, sampling every 200 steps.

Results

The DNA sequences of the mitochondrial 16S, COII, and CR loci, totaling 1040 bp, from 220 individual ticks from four geographic locations yielded 98 concatenated haplotypes whereas the individual loci yielded 13, 35, and 42 haplotypes, respectively (see Tables 2–4 for haplotype distributions among locations and GenBank accession numbers). The number of haplotypes and the haplotype diversity was greatest in the latitudinally intermediate locations (Greenville and Pleasant Valley), intermediate in the southernmost location (Yorktown), and lowest in the northernmost location (Guilderland). This pattern of haplotype diversity was consistent across individual loci and the concatenated dataset (Tables 1, S2). Restriction in gene flow among locations increased with geographical distance (pairwise F_{st} values, Table 5), resulting in pattern of incipient population genetic structure. Share haplotypes among locations also suggest progressive restriction in the gene flow among locations (Table 5), as proximate locations share a greater proportion of haplotypes.

PHYLOGEOGRAPHIC INFERENCE

All phylogeographic models indicate high rates of migration of blacklegged ticks among locations, consistent with the evidence of little population genetic structure (Table 5). Nearly all migration events occurred in the south-to-north direction (38 events). Only one north-to-south migration event was inferred (Figs. 2C, 3). Furthermore, all models strongly implicate the southernmost location as the origin of the clade (root and basal branches of the phylogeny), and thus the source population of the recent range expansion, with high posterior probability support (>75%; Fig. 2A). The more recent colonization of northern locations from populations to the south was also supported by an increase in strongly supported migration events to northern locations over time (Fig. 2). Early migrations events were extremely common to Pleasant Valley (the second most southern location), common to Greenville (the next most southern location), and rare to Guilderland (the northernmost location). Migration events in the more recent past were similarly common to all locations.

Estimates of the time in years of the first migration event into each location also strongly support a progressive south-to-north

Table 3. Frequency of cytochrome oxidase subunit II (COII) haplotypes at each sampling location from each year of collection (2004/2007/2009) and GenBank accession numbers.

Haplotype number	Total	Guilderland	Greenville	Pleasant Valley	Yorktown	Accession numbers
1	44	7/6/4	4/5/6	7/2/0	0/2/1	KR092237
2	38	1/1/4	4/3/5	6/3/6	0/2/3	KR092236
3	37	3/2/1	1/3/1	1/3/5	0/7/10	KR092242
4	12	1/3/0	0/1/1	0/3/2	0/1/0	KR092246
5	10	2/2/0	4/0/0	0/1/1	0/0/0	KR092243
6	10	2/3/5	0/0/0	0/0/0	0/0/0	KR092245
7	8	1/0/0	1/0/0	0/2/0	0/4/0	KR092249
8	5	0/0/0	0/1/1	2/0/0	0/1/0	KR092239
9	5	1/0/0	1/0/0	0/0/1	0/2/0	KR092247
10	4	0/0/0	1/0/2	1/0/0	0/0/0	KR092238
11	4	0/0/0	0/0/0	0/1/3	0/0/0	KR092256
12	4	0/1/2	0/1/0	0/0/0	0/0/0	KR092259
13	3	0/0/0	1/0/0	2/0/0	0/0/0	KR092241
14	3	0/0/0	1/1/0	0/1/0	0/0/0	KR092250
15	3	0/1/0	1/0/0	0/1/0	0/0/0	KR092252
16	3	0/0/0	0/1/0	0/1/1	0/0/0	KR092257
17	3	0/1/0	0/0/0	0/0/0	0/0/2	KR092258
18	3	0/0/2	0/0/1	0/0/0	0/0/0	KR092268
19	2	1/0/1	0/0/0	0/0/0	0/0/0	KR092244
20	2	0/0/0	1/1/0	0/0/0	0/0/0	KR092251
21	2	0/0/1	0/0/0	0/0/0	0/0/1	KR092265
22	2	0/0/0	0/0/1	0/0/0	0/0/1	KR092266
23	1	0/0/0	0/0/0	1/0/0	0/0/0	KR092240
24	1	1/0/0	0/0/0	0/0/0	0/0/0	KR092247
25	1	0/0/0	0/0/0	0/0/0	0/1/0	KR092253
26	1	0/0/0	0/0/0	0/1/0	0/0/0	KR092254
27	1	0/0/0	0/0/0	0/1/0	0/0/0	KR092255
28	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092260
29	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092261
30	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092262
31	1	0/0/0	0/0/0	0/0/0	0/0/1	KR092263
32	1	0/0/0	0/0/0	0/0/0	0/0/1	KR092264
33	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092267
34	1	0/0/0	0/0/1	0/0/0	0/0/0	KR092269
35	1	0/0/0	0/0/1	0/0/0	0/0/0	KR092270

expansion. Ticks in the southernmost location, Yorktown, form the root of the phylogeny which coalesces 57 years ago (34–85 HDP). The first recorded migration event into the second southernmost location, Pleasant Valley, is estimated to have occurred 40 years ago (24–59 HDP) whereas migration events into Greenville (second northernmost location) and Guilderland (northernmost location) are estimated to have occurred 31 (22–49 HDP) and 14 (11–19 HDP) years ago, respectively. Absolute time estimates from these analyses should be interpreted with caution.

Migration among adjacent locations, considered local dispersal (1–100 km), was more common (26 supported migration events) than migration among nonadjacent locations or long-distance dispersal (12 supported migration events; Figs. 2, 3). Most south-to-north long-distance dispersal events with strong posterior support emigrated from the southernmost location and were more frequent to the most proximal nonadjacent location (nine events) than to the northernmost location (two events).

The phylogenetic topologies, the inferred geographic locations of ancestral nodes and the inferred migratory events were robust to the underlying demographic model. The results from all demographic models consistently identified the same migratory events with high posterior probability support. The Extended Bayesian skyline plot (linear model) had the greatest support in both the path sampling and SS algorithms (Table S4). Although only data from the best performing model are presented (Figs. 2, 3), the general phylogeographic inferences and parameter

Table 4. Frequency of control region (CR) haplotypes at each sampling location from each year of collection (2004/2007/2009) and GenBank accession numbers.

Haplotype number	Total	Guilderland	Greenville	Pleasant Valley	Yorktown	Accession numbers
1	35	2/2/1	0/4/1	1/3/5	0/6/10	KR092282
2	31	6/6/2	4/2/5	5/1/0	0/0/0	KR092273
3	21	0/1/0	3/2/3	3/2/3	0/2/2	KR092272
4	17	4/0/2	2/0/0	1/1/0	0/6/1	KR092283
5	16	2/3/3	2/2/2	0/0/1	0/0/1	KR092284
6	14	2/3/5	0/0/0	1/2/1	0/0/0	KR092281
7	13	1/3/0	0/1/1	1/4/2	0/0/0	KR092280
8	11	0/0/0	2/1/1	1/2/1	0/1/2	KR092276
9	6	1/0/2	0/0/2	1/0/0	0/0/0	KR092275
10	5	0/0/0	1/0/0	0/2/0	0/2/0	KR092290
11	4	0/0/0	1/0/2	1/0/0	0/0/0	KR092274
12	4	0/0/0	0/0/1	1/0/0	0/0/2	KR092277
13	3	0/0/1	0/0/0	2/0/0	0/0/0	KR092271
14	3	0/0/0	2/0/0	1/0/0	0/0/0	KR092279
15	3	0/0/0	1/1/0	0/1/0	0/0/0	KR092287
16	3	0/1/1	0/0/0	0/0/0	0/0/1	KR092295
17	2	0/0/0	0/0/0	1/0/1	0/0/0	KR092278
18	2	0/0/0	1/1/0	0/0/0	0/0/0	KR092289
19	2	0/0/0	0/0/0	0/1/0	0/1/0	KR092291
20	2	0/0/0	0/0/0	0/1/0	0/1/0	KR092292
21	2	0/0/0	0/0/1	0/0/0	0/0/1	KR092308
22	1	1/0/0	0/0/0	0/0/0	0/0/0	KR092285
23	1	1/0/0	0/0/0	0/0/0	0/0/0	KR092286
24	1	0/0/0	1/0/0	0/0/0	0/0/0	KR092288
25	1	0/0/0	0/0/0	0/0/0	0/1/0	KR092293
26	1	0/1/0	0/0/0	0/0/0	0/0/0	KR092294
27	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092296
28	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092297
29	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092298
30	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092299
31	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092300
32	1	0/0/0	0/1/0	0/0/0	0/0/0	KR092301
33	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092302
34	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092303
35	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092309
36	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092310
37	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092311
38	1	0/0/0	0/0/0	0/0/1	0/0/0	KR092312
39	1	0/0/1	0/0/0	0/0/0	0/0/0	KR092304
40	1	0/0/1	0/0/0	0/0/0	0/0/0	KR092305
41	1	0/0/1	0/0/0	0/0/0	0/0/0	KR092306
42	1	0/0/0	0/0/1	0/0/0	0/0/0	KR092307

estimates were similar under the majority of the demographic models.

POPULATION DEMOGRAPHY

The demographic reconstructions in Beast (Fig. 2), the demographic estimates derived from the LAMARC algorithm (Fig. 4, Table S3), and the mismatch distribution analyses (Fig. 5) all indicate that the blacklegged tick population size has increased dramatically in the recent past. All demographic models in Beast that allow for population growth produced similar estimates of demographic parameters. The demographic parameters estimated with LAMARC indicate that the population size at each sampled location is increasing, although they differ both in effective population size and growth rate (Fig. 4, Table S3). The

Table 5. Population genetic structure among sampled locations.

	Sampled location			
	Guilderland	Greenville	Pleasant Valley	Yorktown
Guilderland	-	0.03957	0.05905	0.06154
Greenville	9	-	*0.00201*	0.02139
Pleasant Valley	5	11	-	0.01524
Yorktown	7	8	15	-

Pairwise F_{st} values are presented above the diagonal. Statistically significant population differentiation was detected between all location except between Pleasant Valley and Greenville (lack of significance indicated with italics in the table). The number of haplotypes shared among pairs of locations is presented below the diagonal.

demographic estimator, Θ (proportional to the effective population size assuming a constant mutation rate across locations, $\Theta = N_e \times \mu$), indicates that effective population size decreases with increasing latitude (Table S3). The distribution of mismatches found among all sampled ticks (Fig. 5) and among ticks at each location (Fig. S1) differ from the distribution expected in populations experiencing no demographic changes. The unimodal distributions seen in each analysis are commonly interpreted as an increasing population size although they can also indicate increasing geographic range (Rogers and Harpending 1992).

Discussion

Ixodes scapularis ticks were either absent or at densities below detectable levels only a few decades ago in many locations in the Northeastern United States. Over the last decades, tick populations have become present and even prevalent in some previously uninhabited areas. The processes that have generated this dynamic biogeographic pattern may have resulted from either in situ growth of established populations that were maintained at very low densities or by migration and colonization from established populations. We used a Bayesian phylogeographic modeling framework to infer the colonization and demographic history of blacklegged ticks in the Hudson River Valley, New York, a region with documented changes in population densities (Daniels et al. 1998; Khatchikian et al. 2012) and apparent changes in the geographic range of this species. The data and analyses presented indicate that recent migration and colonization of the northern areas of this region has resulted in a rapid range expansion of the tick population (Fig. 4). The range expansion proceeded primarily through progressive and local migration events from southern populations to proximate northern locations, although long-distance migration events were detected. Additionally, all populations show strong signatures of recent and rapid population growth suggesting that both the geographic range and the population densities are increasing.

The apparent increase in the geographic range of blacklegged ticks results not from in situ growth from previously existent populations but from migration of individuals into previously unoccupied territories. Recent population growth from small isolated populations that were previously below detectable levels would cause patterns of strong population genetic structure as seen in previous studies of established *I. scapularis* populations (Qiu et al. 2002; Humphrey et al. 2010; Kelly et al. 2014). The samples from these locations in the Hudson River Valley showed only slight evidence of genetic isolation and incipient population genetic structure (Table 5). A similar pattern can be observed in the distribution of shared haplotypes, suggesting progressive restriction in the gene flow among locations (Table 5). Further, long-standing isolated populations would likely maintain private haplotypes, although no phylogenetic clades appear to originate from the northern locations (Figs. 2, 3). The phylogeographic and population genetic analyses present a pattern of multiple migration events that seeded the populations in areas where ticks were previously undetected (Figs. 2, 3).

Nearly all inferred migration events occur in a south-to-north direction confirming the directionality of the range expansion. Further, reconstruction of the geographic location of ancestral nodes indicated with high confidence that the southernmost location forms the origin of the phylogeny (Fig. 2) and many highly supported emigration events originated from this location (Figs. 2, 3). This conclusion is consistent with decades-old field collections reporting endemic *I. scapularis* in the southern region (Daniels et al. 1998). The reconstruction of the demographic and migratory history of *I. scapularis* in the region suggests that ticks from the southernmost region are the source of migrants that subsequently colonized northern locations.

The majority of strongly supported migratory events occurred between neighboring locations, suggesting local dispersal events were the primary cause of the observed range expansion (Figs. 2, 3). Most of the observed immigration events in the more distant past colonized locations proximal to the southernmost population whereas immigration events in the more recent past colonized all locations equally, generally from the immediate southern neighbor (Fig. 2C). The active dispersal capabilities of ticks are minimal (a few meters) such that immigration necessarily occurs through passive dispersal while attached to a vertebrate host (Leighton et al. 2012). Thus, the dispersal patterns observed suggest that the majority of migration events occur while ticks feed on vertebrate hosts that were not migrating long distances. However, the distances between the locations investigated are considerable for terrestrial vertebrates indicating that migration on multiple different host animals spanning several tick generations

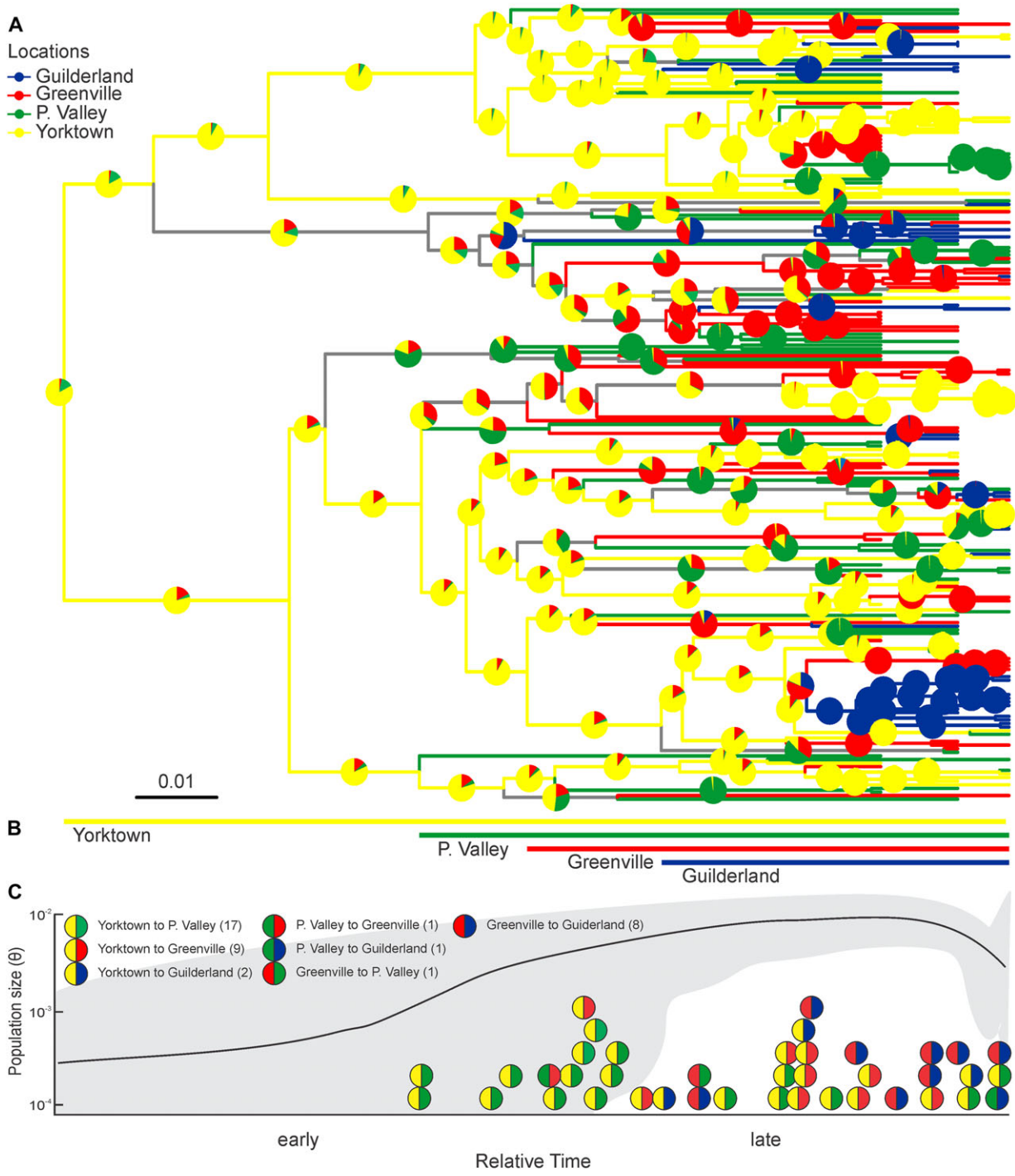

Figure 2. Migratory events occurred primarily in a south-to-north direction among adjacent locations resulting in a gradual *I. scapularis* range expansion. (A) Maximum clade credibility (MCC) tree indicates that the southernmost location is the source population of the recent range expansion with strong posterior probability support across nearly all basal nodes. Further, nearly all observed migration events (31 vs. 1) occurred in a south-to-north direction. The pie charts indicate the Bayesian posterior probability support for the inferred location and the internal branches are coded by location when posterior support >75%. All terminal branches are colored coded according to the location where the sample was collected. Scale bar represent substitutions. (B) Graphical representation of the first observed colonization event with high posterior support (>75%) at each location, indicating a clear south-to-north expansion pattern. (C) Early migration events originated from the southernmost population whereas more recent migration events originated from both the southernmost and the intermediate populations. Colored circles indicate the origin and destination of migration events. The number of migration events between each location is denoted in the parentheses of the legend. The demographic estimates through time are shown as of the effective population size times the mutation rate (Θ; line) and the 95% confidence interval (shaded areas).

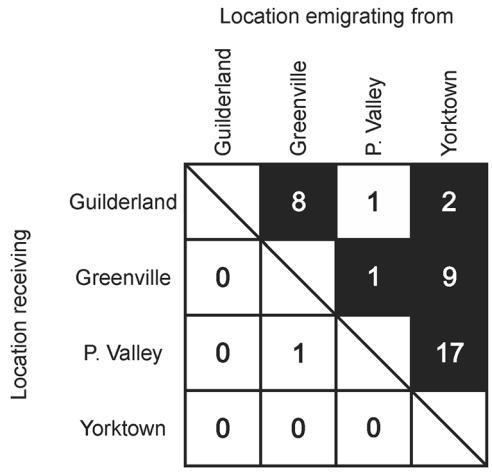

Figure 3. **The majority of the migratory events occurred in a south-to-north direction between adjacent locations. Black boxes indicate Bayesian support (Bayes factors >10) for directional migration between locations calculated using the Bayesian stochastic search selection variable selection (BSSVS). The minimum number of highly supported migratory events necessary to reconcile the phylogeny with the inferred geographic locations is presented in the matrix.**

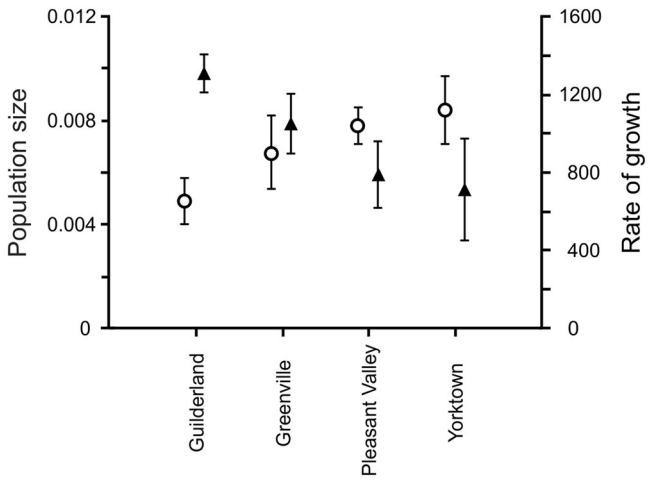

Figure 4. **Effective population sizes are negatively correlated with latitude whereas population growth rates are positively correlated with latitude. The effective population size (N_e) is proportional to Θ (left axis; open circles) assuming a constant mutation rate across locations ($\Theta = N_e \times \mu$). The population growth parameter (g) (right axis; triangles) represents the change in the effective population size over time ($\Theta_t = \Theta_{present} \times \exp(-gt)$).**

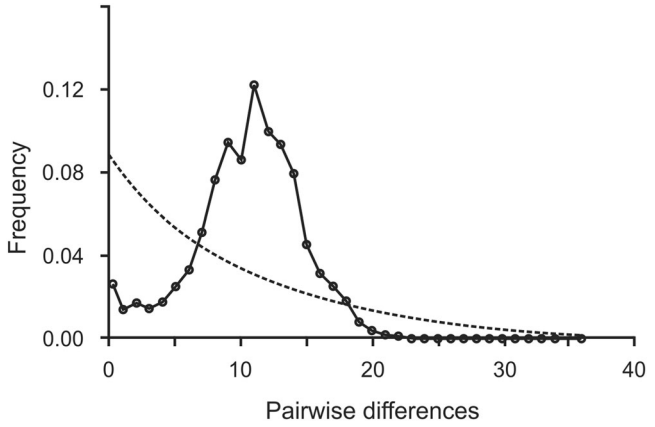

Figure 5. **The observed frequency distribution of pairwise differences among haplotypes (solid line) differs from that expected in populations that are not expanding (dashed line). The observed unimodal frequency distribution is most commonly interpreted as an indication of increasing population sizes.**

may be necessary to migrate between the sampled locations. Future phylogeographic studies at finer geographic scales are needed to assess the role different vertebrates play in the observed gradual range expansion. The less-frequent long-distance dispersal events (Figs. 2, 3), however, very likely occurred on migratory birds as they can traverse very large geographic distances in the few days that ticks remain attached while blood feeding (see Madhav et al. 2004; Ogden et al. 2006, 2008b, 2009; Elias et al. 2011; Leighton et al. 2012). The data and analyses presented are compatible and complementary to previous studies that identified migratory birds as the vehicle for long-distance migration into Canada and some islands (e.g., Ogden et al. 2008b; Leighton et al. 2012).

The primary conclusions about the direction, rate, and relative timing of migration in this study rely only on the relative timing of events although the absolute time of colonization of each location can also be estimated from the coalescent analysis (Lemey et al. 2009b). Estimates of absolute time suggest that the most northern location was colonized 14 years ago, the second most northern 31 years ago, and the third most northern location 40 years ago. Thus, the migratory dynamics in this geographic region occurred on a recent (<100 years) as opposed to a glacial (>1000 years) time scale. These estimates are supported by field collections that found no evidence of tick populations in the northern parts of the study area in the early 2000s, despite robust sampling (Daniels et al. 1998). However, absolute times should be interpreted with extreme caution. Reconstruction of absolute time from a phylogeny is contentious and relies on assumptions that cannot be verified in this study.

Demographic reconstructions using multiple algorithms as well as the mismatch distribution analyses consistently indicated that *I. scapularis* densities recently increased throughout the region (Figs. 2, 4, 5 and Table S3). Furthermore, the estimated demographic parameters for each sampled location (Fig. 4, Table S3) confirm that current population growth rates are positively correlated with latitude whereas effective population size estimates decrease with increasing latitude. The pattern of these results is typical of a directional range expansion. Thus, the northern locations, which were recently colonized by migrants from the south,

have smaller effective population sizes but are experiencing rapid population growth as expected in the initial phases of growth of recently established populations (Andrewartha and Birch 1954). The southern locations represent longer standing populations with lower current growth rates as expected of populations approaching stable densities (Andrewartha and Birch 1954).

The most parsimonious model of population and geographic range expansion indicates that previously uncolonized areas were colonized primarily by emigrants from proximal southern populations. As the number of blacklegged ticks increased in these newly colonized locations, the new locations became the source of emigrants to proximal, uninhabited areas (Fig. 2), explaining the apparent progressive increase in immigrants arriving to northern locations in the recent past. Long-distance migration events are responsible for a smaller proportion of the gene flow, although the rate of long-distance migration is not great enough to homogenize the tick population in our study area (Table 5). The combined effect of these two dispersal modes produced the complex phylogeographical signature observed in our dataset.

The inferred demographic history and the reconstruction of the phylogeographic patterns among locations suggest a recent and rapid population and geographic range expansion. It is important to note that migrants entering our samples from areas outside of our sampling locations, which undoubtedly occurred, may be responsible for the high overall genetic diversity detected in this study (as exemplified by the high number of haplotypes detected). However, the analyses employed are robust to such samples (see Lemey et al. 2009a; Bloomquist et al. 2010; Molak et al. 2012) and can effectively infer the phylogeographic history of *I. scapularis.* Phylogeographic analysis of mitochondrial markers, which reflects the history of female ticks and whose advantages and disadvantages have been extensively discussed (see Avise 2004), may be a limitation of this study. However, this limitation has no qualitative or quantitative impact on the demographic or phylogeographic inferences in the current study.

The population and range expansion of *I. scapularis* ticks into novel areas was driven either by adaptation to the novel local environments or by recent changes in the environment itself. Although the presented data show no evidence of phenotypic adaptation, adaptive responses to environmental factors like those reported in other arthropod species (e.g., Bradshaw and Holzapfel 2001; Khatchikian et al. 2010) could have facilitated the observed range expansion. Alternatively, the rapid geographic and population expansion may have been facilitated by the changes in landscape and climate factors, both of which have changed in recent decades (Estrada-Peña 2002; Ostfeld et al. 2006; Ogden et al. 2008a; Ostfeld 2009). If environmental factors caused the recent *I. scapularis* population dynamics, fine-scale studies may identify specific sets of environmental factors that permit migration, colonization, and population growth. Regardless of the cause of the observed population dynamics, the recent establishment of *I. scapularis* populations in novel areas has increased the suitability of these areas for tick-borne pathogens which has resulted in increased disease transmission to humans (e.g., Hamer et al. 2010; Khatchikian et al. 2012; Kelly et al. 2014). Further, the continuous influx of migrants also reduces the chance of local extinctions of tick and pathogen populations and increases the chances of their regional persistence. An important evolutionary implication of migration is the introduction of genetic variation into novel selective regimes that has the potential to affect future population dynamics and pathogen transmission.

These results demonstrate that short-distance migration into uninhabited areas, resulting in a geographic range expansion, has driven recent evolutionary dynamics in *I. scapularis* populations in the Northeastern United States. Migration of *I. scapularis* individuals to nearby locations has resulted in biological evolution that altered the allele frequencies and may have introduced genetic variants into geographic areas with novel selective environments. The fine temporal and spatial scale of samples analyzed allowed for precise estimates of the rate, timing, and direction of individual migratory events and an accurate understanding of the process of migration that caused the observed patterns of genetic variation. The resulting comprehension of the fine-scale migratory process is essential to interpret patterns of genetic variation across broad geographic regions that may have been isolated for long periods of time (see also Brinkerhoff et al. 2010; Humphrey et al. 2010; Kelly et al. 2014). These results could be valuable to future studies aimed at identifying mechanistic drivers of migration which could be targeted to improve public health (Kurtenbach et al. 2006; Khatchikian et al. 2009, 2011; Brisson et al. 2010; Jaenson and Lindgren 2011).

ACKNOWLEDGMENTS

The authors would like to express their sincere gratitude to many public and private landholders for granting us property access to conduct this research. We would also like to thank J. Kokas, R. Falco, S. Kogut, J. H. Lee, M. VanDeusen, J. Hallisey, and many Entomological Assistants, student interns, and county health department staff for their assistance in collection and identification of ticks. The authors are very grateful to the many anonymous reviewers who helped to improve this manuscript. This work was supported in part by grants from the National Institutes of Health (AI076342 and AI097137) and from the Burroughs Wellcome Fund (1012376).

DATA ARCHIVING

DNA sequences: GenBank accession numbers for 16S KR092223–KR092235, for COII KR092236–KR092270, and for CR KR092271–KR092312.

LITERATURE CITED

Andrewartha, H. G., and L. C. Birch. 1954. The distribution and abundance of animals. University Chicago Press, Chicago, IL.

Avise, J. C. 2004. Molecular markers, natural history, and evolution. Sinauer Associates, Inc., Sunderland, MA.

Baele, G., P. Lemey, T. Bedford, A. Rambaut, M. A. Suchard, and A. V. Alekseyenko. 2012. Improving the accuracy of demographic and molecular clock model comparison while accommodating phylogenetic uncertainty. Mol. Biol. Evol. 29:2157–2167.

Bloomquist, E. W., P. Lemey, and M. A. Suchard. 2010. Three roads diverged? Routes to phylogeographic inference. Trends Ecol. Evol. 25:626–632.

Bradshaw, W. E., and C. M. Holzapfel. 2001. Genetic shift in photoperiodic response correlated with global warming. Proc. Natl. Acad. Sci. USA 98:14509–14511.

Brinkerhoff, R. J., S. J. Bent, C. M. Folsom-O'Keefe, K. Tsao, A. G. Hoen, A. G. Barbour, and M. A. Diuk-Wasser. 2010. Genotypic diversity of *Borrelia burgdorferi* strains detected in *Ixodes scapularis* larvae collected from North American songbirds. Appl. Environ. Microbiol. 76:8265–8268.

Brisson, D., M. F. Vandermause, J. K. Meece, K. D. Reed, and D. E. Dykhuizen. 2010. Evolution of northeastern and midwestern *Borrelia burgdorferi*, United States. Emerging Infect. Dis. 16:911–917.

Casati, S., M. V. Bernasconi, L. Gern, and J. C. Piffaretti. 2008. Assessment of intraspecific mtDNA variability of European *Ixodes ricinus* sensu stricto (Acari: Ixodidae). Infect. Genet. Evol. 8:152–158.

Daniels, T. J., T. M. Boccia, S. Varde, J. Marcus, J. Le, D. J. Bucher, R. C. Falco, and I. Schwartz. 1998. Geographic risk for Lyme disease and human granulocytic ehrlichiosis in southern New York state. Appl. Environ. Microbiol. 64:4663–4669.

Drummond, A. J., M. Kearse, J. Heled, and R. Moir. 2009. Geneious v4.8.

Drummond, A. J., M. A. Suchard, D. Xie, and A. Rambaut. 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29:1969–1973.

Elias, S. P., R. P. Smith, S. R. Morris, P. W. Rand, C. Lubelczyk, and E. H. Lacombe. 2011. Density of *Ixodes scapularis* ticks on Monhegan Island after complete deer removal: a question of avian importation? J. Vector Ecol. 36:11–23.

Estrada-Peña, A. 2002. Increasing habitat suitability in the United States for the tick that transmits Lyme disease: a remote sensing approach. Environ. Health Perspect. 110:635–640.

Excoffier, L., and H. E. L. Lischer. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol. Ecol. Res. 10:564–567.

Ferreira, M. A. R., and M. A. Suchard. 2008. Bayesian analysis of elapsed times in continuous-time Markov chains. Can. J. Stat. 36:355–368.

Firth, C., A. Kitchen, B. Shapiro, M. A. Suchard, E. C. Holmes, and A. Rambaut. 2010. Using time-structured data to estimate evolutionary rates of double-stranded DNA viruses. Mol. Biol. Evol. 27:2038–2051.

Foley, E. A., C. E. Khatchikian, J. Hwang, J. Ancca-Juarez, K. Borrini-Mayori, V. R. Quispe-Machaca, M. Z. Levy, and D. Brisson. 2013. Population structure of the Chagas disease vector, *Triatoma infestans*, at the urban-rural interface. Mol. Ecol. 22:5162–5171.

Ginsberg, H. S., and C. P. Ewing. 1989. Comparison of flagging, walking, trapping, and collecting from hosts as sampling methods for northern deer ticks, *Ixodes dammini*, and lonestar ticks, *Amblyomma americanum* (Acari: Ixodidae). Exp. Appl. Acarol. 7:313–322.

Griffiths, R. C., and S. Tavare. 1994. Sampling theory for neutral alleles in a varying environment. Philos. Trans. R. Soc. B 344:403–410.

Hamer, S. A., J. I. Tsao, E. D. Walker, and G. J. Hickling. 2010. Invasion of the Lyme disease vector *Ixodes scapularis*: implications for *Borrelia burgdorferi* endemicity. Ecohealth 7:47–63.

Heled, J., and A. Drummond. 2008. Bayesian inference of population size history from multiple loci. BMC Evol. Biol. 8:289.

Humphrey, P. T., D. A. Caporale, and D. Brisson. 2010. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. Evolution 64:2653–2663.

Jaenson, T. G. T., and E. Lindgren. 2011. The range of *Ixodes ricinus* and the risk of contracting Lyme borreliosis will increase northwards when the vegetation period becomes longer. Ticks Tick Borne Dis. 2:44–49.

Kaplan, L., D. Kendell, D. Robertson, T. Livdahl, and C. Khatchikian. 2010. *Aedes aegypti* and *Aedes albopictus* in Bermuda: extinction, invasion, invasion and extinction. Biol. Invasions 12:3277–3288.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., Buxton, S., Cooper, A., Markowitz, S., Duran, C., Thierer, T., Ashton, B., Mentjies, P., and Drummond, A. 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics 28:1647–1649.

Kelly, R. R., D. Gaines, W. F. Gilliam, and R. J. Brinkerhoff. 2014. Population genetic structure of the Lyme disease vector *Ixodes scapularis* at an apparent spatial expansion front. Infect. Genet. Evol. 27:543–550.

Khatchikian, C., J. Dennehy, C. Vitek, and T. Livdahl. 2010. Environmental effects on bet hedging in *Aedes* mosquito egg hatch. Evol. Ecol. 24:1159–1169.

Khatchikian, C., F. Sangermano, D. Kendell, and T. Livdahl. 2011. Evaluation of species distribution model algorithms for fine-scale container-breeding mosquito risk prediction. Med. Vet. Entomol. 25:268–275.

Khatchikian, C. E., J. J. Dennehy, C. J. Vitek, and T. Livdahl. 2009. Climate and geographic trends in hatch delay of the treehole mosquito, *Aedes triseriatus* Say (Diptera: Culicidae). J. Vector Ecol. 34:119–128.

Khatchikian, C. E., M. Prusinski, M. Stone, P. B. Backenson, I.-N. Wang, M. Z. Levy, and D. Brisson. 2012. Geographical and environmental factors driving the increase in the Lyme disease vector *Ixodes scapularis*. Ecosphere 3:art85.

Khatchikian, C. E., E. A. Foley, C. M. Barbu, J. Hwang, J. Ancca-Juárez, K. Borrini-Mayori, V. R. Quıspe-Machaca, C. Naquira, D. Brisson, and M. Z. Levy. 2015. Population structure of the Chagas disease vector *Triatoma infestans* in an urban environment. PloS Negl. Trop. Dis. 9:e0003425.

Kingman, J. F. C. 1982. On the genealogy of large populations. J. App. Prob. 19:27–43.

Kuhner, M. K. 2006. LAMARC 2.0: maximum likelihood and Bayesian estimation of population parameters. Bioinformatics 22:768–770.

Kurtenbach, K., K. Hanincova, J. I. Tsao, G. Margos, D. Fish, and N. H. Ogden. 2006. Fundamental processes in the evolutionary ecology of Lyme borreliosis. Nat. Rev. Microbiol. 4:660–669.

Lartillot, N. and H. Philippe. 2006. Computing Bayes factors using thermodynamic integration. Sys. Biol. 55:195–207.

Leberg, P. L. 2002. Estimating allelic richness: effects of sample size and bottlenecks. Mol. Ecol. 11:2445–2449.

Leighton, P. A., J. K. Koffi, Y. Pelcat, L. R. Lindsay, and N. H. Ogden. 2012. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. J. App. Ecol. 49:457–464.

Lemey, P., A. Rambaut, A. J. Drummond, and M. A. Suchard. 2009a. Bayesian phylogeography finds its roots. PloS Comput. Biol. 5:e1000520.

Lemey, P., M. Suchard, and A. Rambaut. 2009b. Reconstructing the initial global spread of a human influenza pandemic: a Bayesian spatial-temporal model for the global spread of H1N1. PLoS Curr. Infl. 1:RRN1031.

Librado, P., and J. Rozas. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452.

Lounibos, L. P. 2002. Invasions by insect vectors of human disease. Annu. Rev. Entomol. 47:233–266.

Madhav, N. K., J. S. Brownstein, J. I. Tsao, and D. Fish. 2004. A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). J. Med. Entomol. 41:842–852.

Minin, V. N., E. W. Bloomquist, and M. A. Suchard. 2008. Smooth skyride through a rough skyline: Bayesian coalescent-based inference of population dynamics. Mol. Biol. Evol. 25:1459–1471.

Molak, M., E. D. Lorenzen, B. Shapiro, and S. Y. W. Ho. 2012. Phylogenetic estimation of timescales using ancient DNA: the effects of temporal sampling scheme and uncertainty in sample ages. Mol. Biol. Evol. 30:253–262.

Ogden, N. H., A. Maarouf, I. K. Barker, M. Bigras-Poulin, L. R. Lindsay, M. G. Morshed, J. O'Callaghan C, F. Ramay, D. Waltner-Toews, and D. F. Charron. 2006. Climate change and the potential for range expansion of the Lyme disease vector *Ixodes scapularis* in Canada. Int. J. Parasitol. 36:63–70.

Ogden, N. H., M. Bigras-Poulin, K. Hanincova, A. Maarouf, C. J. O'Callaghan, and K. Kurtenbach. 2008a. Projected effects of climate change on tick phenology and fitness of pathogens transmitted by the North American tick *Ixodes scapularis*. J. Theor. Biol. 254:621–632.

Ogden, N. H., L. R. Lindsay, K. Hanincova, I. K. Barker, M. Bigras-Poulin, D. F. Charron, A. Heagy, C. M. Francis, C. J. O'Callaghan, I. Schwartz, et al. 2008b. 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. 74:1780–1790.

Ogden, N. H., L. R. Lindsay, M. Morshed, P. N. Sockett, and H. Artsob. 2009. The emergence of Lyme disease in Canada. Can. Med. Assoc. J. 180:1221–1224.

Ostfeld, R. S. 2009. Climate change and the distribution and intensity of infectious diseases. Ecology 90:903–905.

Ostfeld, R. S., C. D. Canham, K. Oggenfuss, R. J. Winchcombe, and F. Keesing. 2006. Climate, deer, rodents, and acorns as determinants of variation in Lyme-disease risk. PLoS Biol. 4:e145.

Qiu, W. G., D. E. Dykhuizen, M. S. Acosta, and B. J. Luft. 2002. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. Genetics 160:833–849.

Rambaut, A., M. A. Suchard, D. Xie, and A. J. Drummond. 2014. Tracer v1.6. Available at http://beast.bio.ed.ac.uk/Tracer.

Rogers, A. R., and H. Harpending. 1992. Population growth makes waves in the distribution of pairwise genetic differences. Mol. Biol. Evol. 9:552–569.

Simon, C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. 87:651–701.

Suchard, M., R. Weiss, and J. S. Sinsheimer. 2005. Models for estimating Bayes factors with application to phylogeny and tests of monophyly. Biometrics 61:665–673.

Xie, W., P. O. Lewis, Y. Fan, L. Kuo, and M.-H. Chen. 2011. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. Syst. Biol. 60:150–160.

Associate Editor: A.-L. Laine
Handling Editor: R. Shaw

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1. PCR amplification.

Appendix S2. Complete list of genetic diversity indices.

Appendix S3. Demographic parameters.

Appendix S4. Path sampling (PS) and stone sampling (SS) values.

Appendix S5. Assessment of the robustness of the phylogeographical signal and inferences in our sampling scheme using a random operator that shuffles the location for each tip. A standard Beast analysis is run with the addition of the operator that randomizes the tip location. Once that the MCMC is finished, all nodes are annotated using the posterior distribution present in the trees. As the operator only swaps locations it does not change the original number of tips per location. We run 20 analyses and verified the even distribution of the root location (and main branches of the tree).

Appendix S6. Frequency distributions of pairwise differences.

Figure S6. Observed frequency distributions of pairwise differences among haplotypes (solid line) within each location. In all cases, the observed patterns differ from that expected in populations that are not expanding (dashed line).

Table S1. Primers used for *mt*DNA amplification of the 16S, the cytochrome oxidase subunit II (COII) and the control region (CR).

Table S2. Genetic diversity indices including the number of haplotypes (H), the standardized number of haplotypes (H'), haplotype diversity (H_d,), and standardized haplotype diversity (H_d') for each sampled location. The 95% confidence interval presented in parentheses.

Table S3. Demographic parameters Θ and g estimated for the populations at each sampled location. The effective population size (N_e) is proportional to Θ assuming a constant mutation rate across locations ($\Theta = N_e{}^{*}\mu$). The population growth parameter (g) is equivalent to the change in effective population size over time ($\Theta t = \Theta_{present}{}^{*} \exp(-gt)$).

Table S4. Path sampling (PS) and stone sampling (SS) values obtained for each different demographic model evaluated.