Fall 2015

Genetic Structure of Ixodes Scapularis Say 1821 (Acari: Ixodidae), The Blacklegged Tick, By Microsatellite Analysis

John Ludwig

GENETIC STRUCTURE OF *IXODES SCAPULARIS* SAY 1821 (ACARI: IXODIDAE), THE BLACKLEGGED TICK, BY MICROSATELLITE ANALYSIS.

by

JOHN LUDWIG

(Under the Direction of Lorenza Beati)

ABSTRACT

The blacklegged tick, *Ixodes scapularis* (Acari: Ixodidae) is the subject of intense research due to its economic importance from being a vector of several diseases. Previous studies have examined its taxonomic status and more recently, its genetic structure. This study embarked upon resolving the questions surrounding the genetic structuring of the tick, by using different molecular markers. Molecular data were first phylogenetically analyzed by using four different mitochondrial and nuclear markers (12SrDNA n=483, Dloop n=432, ITS2 n=156, Actin n=117). A Bayesian tree was inferred from the concatenated mitochondrial dataset, whereas nuclear gene markers were found to be uninformative. Nine microsatellites loci (n=462) were also genotyped and used to study the population structure of the same individual tick samples used for phylogenetic analysis. Information generated with the different markers was compared. As other studies have reported, the Bayesian analysis identified six clades, all with strong support. Microsatellite analysis showed that *I. scapularis* is genetically structured at the "state" level, and that genetic distance increases with geographic distance ($r = 0.582$, $p = 0.001$; $r = 0.704$ $p = 0.005$). However, it found no correlation between distances generated from microsatellites and mtDNA sequences ($r = 0.025$ $p = 0.0964$), further supporting the opinion that mitochondrial gene sequences should be used cautiously when delimiting species as they can be mere witnesses of past vicariant events that did not result in speciation. Our work indicates that microsatellite markers, codominant bi-parentally inherited markers, can elucidate long-standing questions better than other markers.

Index words: *Ixodes scapularis*, Microsatellites, Mitochondrial genes, Nuclear Gene, Population Genetics, Phylogeny

GENETIC STRUCTURE OF *IXODES SCAPULARIS* SAY 1821 (ACARI: IXODIDAE), THE BLACKLEGGED TICK, BY MICROSATELLITE ANALYSIS.

by

JOHN LUDWIG

B.S., Georgia Southern University, 2013

A Thesis Submitted to the Graduate Faculty of Georgia Southern University in Partial Fulfillment of the Requirements for the Degree

Master of Science

Statesboro, Georgia

GENETIC STRUCTURE OF *IXODES SCAPULARIS* SAY 1821 (ACARI: IXODIDAE), THE BLACKLEGGED TICK, BY MICROSATELLITE ANALYSIS.

by

JOHN LUDWIG

Major Professor: Lorenza Beati

Committee: Laura Reggassa

Lance Durden

Electronic Version Approved:
December 2015

TABLE OF CONTENTS

Chapter 1.

Introduction....8

Material and Methods....10

Results....13

Discussion....15

Conclusion....17

References....31

INTRODUCTION

Ixodes scapularis Say 1821 (Acari: Ixodidae) is a hard tick that occurs in the Eastern half of the U.S., and parts of Mexico and Canada (Keirans et al. 1996). Since the discovery of the role of *I. scapularis* in the transmission of Lyme disease (Burgdorfer et al. 1982; Steer et al. 2004), this tick's ecology, biology, and association with pathogens have been the object of intense scrutiny. Similarly, because of the economic and public health impact of the tick as vector of pathogens (Meltzer et al. 1999; Zhang et al. 2006; Hu 2012; Stanek et al. 2012), its taxonomic status and, later, genetic diversity have been examined and reassessed repeatedly.

Spielman et al. (1979) divided *I. scapularis,* based on morphological characters, into two species with *I. scapularis* occupying the Southern half of the Eastern U.S. and *I. dammini* the Northern half. Oliver et al. (1993) relied on reciprocal crosses, karyotype, and isozyme data to relegate *I. dammini* to a junior synonym of *I. scapularis*. Wesson et al. (1993) and McLain et al. (1995) analyzed samples from seven different states, by using two nuclear markers, the ribosomal internal transcribed spacers, ITS1 and ITS2, to determine the taxonomic status of *I. scapularis.* The results of their analyses confirmed that *I. dammini* should be treated as a synonym of *I. scapularis* because the ITS genotypes did not cluster into geographically cohesive groups. Rich et al. (1997), nevertheless, analyzed an ITS2 dataset and demonstrated that ITS sequences are not phylogenetically sound markers, because copies of the gene within a single tick might be more diverse than copies from different ticks. However, their criticism was mostly overseen, because their analysis was concentrated on a very short portion (289 bp) of the ITS2 sequence, which was not considered to be representative of the full length of the gene. Norris et al. (1996) used the ribosomal mitochondrial markers 16SrDNA and 12SrDNA sequenced from ticks from fifteen states for a phylogenetic analysis. They identified two major lineages, called the southern clade and the American clade. As the American clade was found to include specimens from northern and southern latitudes, the study provided additional evidence for synonymizing *I. dammini* with *I. scapularis.* Qiu et al. (2002) sequenced 16SrDNA mitochondrial fragments of *I. scapularis* from nine different states along the Atlantic coast of the U.S. Their minimum spanning network also identified two major lineages, one star-shaped and of recent origin corresponding to the American clade, and a more complex cluster corresponding to the southern clade. Humphrey et al. (2010) and Sakamoto et al. (2014) came to similar

conclusions and also showed that there is little variability within a larger number of nuclear genes throughout the distributional range of *I. scapularis*. As an alternative to ribosomal nuclear gene markers, the nuclear protein-coding gene actin 1 has been investigated to expand the number of available nuclear markers able to further resolve arachnid, and tick, phylogenetics (da Silva Vaz et al. 2005; Vink et al. 2008).

Sequencing of the genome of *I. scapularis* (Van Zee et al. 2007) provided the scientific community with novel tools. Van Zee et al. (2013) selected 9 SNP (single nucleotide polymorphic) loci within nuclear gene sequences, for examining their prevalence and diversity in a total of 40 *I. scapularis* from 4 representative states. Although their sampling was relatively limited, their SNP analysis confirmed the occurrence of genetically more variable populations in the South when compared with ticks from the North.

Earlier reports on microsatellites (short tandem repeats) in *I. scapularis* stated that these were relatively rare in this tick genome, when compared to other arthropod species (Fagerberg et al. 2001; Van Zee et al. 2013). Therefore, they were not considered to be interesting markers for the study of the genetic structure of *I. scapularis* and were mostly ignored by the scientific community. Nevertheless, the analysis of microsatellite loci has proven to be very informative for several other tick species, including *I. ricinus*, a close European relative of *I. scapularis* (Delaye et al. 1998; de Meeûs et al. 2002, 2004; Kempf et al. 2010). By analyzing microsatellite loci, these authors investigated topics as different as genetic structure, sex-biased tick dispersal, and host-race formation. Rosenthal and Spielman (2004) used one of the *I. ricinus* microsatellite markers (IR27) to demonstrate that the Northern and Southern populations of *I. scapularis* are genetically isolated. Microsatellite markers are, so far, considered to be the best tool for the study of population genetics (Lehmann et al. 1996; Goldstein and Schlötterer 1999; de Meeûs 2011). Nevertheless, population genetic analyses based on a single microsatellite locus might be misleading.

The main goals of the present study were: to expand the geographical range of earlier analyses by including samples from the whole distribution range of *I. scapularis*; to analyze the exact same samples by using and comparing different genetic markers, mitochondrial and nuclear gene sequences, and 9 microsatellite loci.

MATERAILS AND METHODS

Sampling

Ixodes scapularis samples were obtained from collection sites (Figure 1) by two methods: vegetation flagging and host survey. Host sampling was conducted at a deer check station, (David W. Force Park, Ellicott City, MD) with ticks being removed directly from deer. Other samples were obtained from colleagues, and members of the Lyme Disease Gradient Consortium (Tables 1 and 2). Collected ticks were stored in 70% ethanol until used for DNA extraction. Ticks were identified using taxonomic keys for *Ixodes* (Clifford et al. 1961; Keirans and Clifford 1978; Durden and Keirans 1996; Keirans et al. 1996).

DNA extraction

DNA was extracted following previously established methods (Beati and Keirans 2001; Beati et al. 2012), using Qiagen DNeasy Blood and Tissue Kits(Qiagen, Chatworth, CA) with proteinase K (Roche, Indianapolis, IN). By excising a small portion of the tick idiosoma before the lysis step, the exoskeleton of each tick could be preserved as a voucher specimen. DNA was stored at 4°C.

PCR amplification and sequencing

Samples were amplified for two mitochondrial gene markers, 12SrRNA (small ribosomal subunit RNA) and D-loop or control region (DL), the nuclear gene markers ITS2 (internal transcribed spacer) and Actin (protein encoding, microfilaments). Samples were also amplified for nine microsatellite loci: L1 (Fagerberg et al. 2001), IR27 (Delaye et al. 1998), and 7 additional loci (L3, L11, L15, L16, L17, L18, and L19) selected in our laboratory from the *I. scapularis* genomic bac files in GenBank (Beati unpublished data). (Primers are listed in Table 3 and 4, PCR mix in table 5, and amplification conditions are listed in table 6 and 7.) For microsatellite PCRs, the forward primer was fluorescently labeled (Life technologies, Carlsbad, CA). Amplicons were sent to the High Throughput Sequencing facility at The University of Washington for purifying and sequencing, while the microsatellite amplicons were sent to the DNA Analysis Facility at Yale University for genotyping.

Cloning

In order to compare the diversity of the full length of the ITS2 gene sequences within single tick specimens and between ticks from different populations, ITS2 amplicons were cloned using the TOPO TA cloning kit (Invitrogen, Carlsbad, CA) following the manufacturer's protocol. Five *E. coli* colonies were selected randomly for each analyzed tick and their inserts were amplified with M13 primers, and subsequently purified and sequenced at the High Throughput Sequencing facility at The University of Washington. The complementary strands of each amplicon were assembled with Sequencher V5.4 (Gene Codes 2013). The assembled sequences were aligned by using MacClade V4.08 (Maddison and Maddison 2005). Secondary structure was considered in aligning 12SrDNA (Beati and Keirans 2001) and DL (Zhang and Hewitt 1997) and gaps created by a single sequence were eliminated. DNASP V5.10 (Librado and Rozas 2009) was used to identify unique haplotypes. Aligned datasets of unique 12SrDNA and D-loop sequences were concatenated in a single data matrix. The ITS2 dataset was visually examined and divergence values between sequences from the same tick individuals and between sequences from different samples were compared. Actin sequences were aligned and compared for genetic diversity. Sequences will be submitted to GenBank and accession numbers listed.

Phylogenetic analysis

A Bayesian tree was generated with Mr. Bayes V3.2.5 (Ronquist and Huelsenbeck 2003) by analyzing the concatenated mitochondrial dataset (n=289 unique sequences). Analyses were conducted twice, each with four chains (1,000,000 generations). The number of reiterations was increased until a standard deviation of split frequencies of < 0.01 was obtained. Trees were sampled every 100^{th} repetition and only those obtained after the 0.01 standard deviation threshold was reached were kept. The 50% majority-rule consensus tree of the remaining trees was inferred and posterior probabilities were recorded for each branch. The outgroup species was *Ixodes affinis* Neumann 1989, a tick belonging, like *I. scapularis*, to the subgenus *Ixodes* and the *Ixodes ricinus* L. complex (Kerians et al. 1996; Xu et al. 2003)

Genotyping

Allele diversity for each microsatellite marker was scored with Genemarker V2.4.0 (Softgenetics 2012). Data were recorded in a spreadsheet and the program CREATE V1.37

(Coombs et al. 2008) was used to convert the data into formats compatible for several population genetics programs.

Comparison between mitochondrial and microsatellite markers

Genetic differences between identified mitochondrial clades were investigated using Wright's F_{st} (Wright 1965) with Weir and Cockerham's unbiased estimator (Weir and Cockerham 1984) (10,000 permutations) in Fstat V2.9.4 (Goudet 2003). For this test, adults were grouped based on state, clade and cohort (a cohort being defined as a group of ticks active during the same year and season). Because of the occurrence of null alleles (see below), FreeNA (Chapuis and Estoup 2007) was used to correct F_{st} values. The dataset was transformed, with missing data recoded as new alleles (999). Because multiple non-independent tests were performed, a sequential Bonferroni (Holm 1979) was used to correct the data set and the program R V3.2.0 (R-Foundation for Statistical Computing 2015) was used to conduct a unilateral binomial test to check for significance.

In addition, a matrix of maximum likelihood genetic distances for the 12SrDNA dataset was generated by using PAUP*4 Beta 10 (Swofford 2015). A corresponding matrix of Cavalli-Sforza and Edward's chord microsatellite distances (Cavalli-Sforza and Edwards 1967) was generated with MSA (Dieringer and Schlötterer 2003). A mantel test was performed to determine correlation between the two matrices by using the online program zt (Bonnet and Van de Peer 2002), which could handle the large size of our combined matrices (462 columns and rows). The correlation (*r*) between the matrices was computed with one of the two matrices randomized 1,000,000 times.

Genetic structure

To determine at which level (state or region) the populations are significantly subdivided, only data from adult ticks (n=285) were analyzed with program R and the addition of the package HierFstat V0.04-10 (Goudet 2005). The collection sites were subdivided by ecologically distinct regions (Figure 1) and by state. Two hierarchical F-statistics were measured: F_{sr}, the inbreeding in states relative to inbreeding in regions and $F_{rt,}$ inbreeding in region relative to inbreeding in the total sample. The significant deviation from 0 was tested by

1,000 randomizations of individuals between states within each region for F_{sr} and by 1000 randomizations of states between regions within total data set for F_{rt}.

Fstat was used for multiple tests during analysis. First, it was used to test linkage disequilibrium, a nonrandom association of alleles, within states at each locus (1000 permutations). Because there were as many tests as locus pairs (36), an exact binomial unilateral test was conducted. To identify individual significant results, a sequential Bonferroni procedure was performed.

Deviation from Hardy-Weinberg was evaluated by using Weir and Cockerham's (Weir and Cockerham 1984) unbiased estimator of Wright's F_{is} (Wright 1965) in Fstat. A total of 10,000 randomizations of alleles between individuals within each subsample were calculated and confidence intervals around the mean calculated by jackniving over standard errors.

When a mutation in the primer binding site causes an allele to fail to amplify, null alleles are recorded, even though the microsatellite might be present. The occurrence of null alleles was inspected with Microchecker V2.2.3 (Van Oosterhout et al. 2004), and their frequencies estimated with Brooksfield's 2^{nd} method (Brookfield 1996). A unilateral exact binominal test, to check goodness of fit between expected and observed null alleles at each locus per subsample, was carried out in R.

In order to test whether tick populations are isolated by distance, we used the only adult tick cohorts represented by several geographical sites (cohorts C6 with 121 ticks, and C7 with 32 ticks). Ticks were sampled from October 2010 to April 2011 for C6 and in Fall 2012 for Cohort C7. Isolation by distance was tested separately in C6 and C7 with a Mantel test based on Cavalli-Sforza and Edward's chord distance (Cavalli-Sforza and Edwards 1967) corrected for null alleles by FreeNa (Chapuis and Estoup 2007). The UTM geographical coordinates of each site were used to compute geographic distances with Genepop 4 (Rousset 2008). Chord distance matrix corrected for null alleles indeed represents the most powerful option (Séré et al., unpublished data). For the Mantel test, we randomized one of two matrices 10,000 times with Genepop on the Web (Rousset 2008; Morgan 2013; Raymond and Rousset 2013).

RESULTS

Sequencing and phylogenetic analyses

Mitochondrial genes:

A total of 483 12SrDNA sequences (361 bp) and 432 DL sequences (516 bp) were generated yielding 166 and 223 unique haplotypes, respectively. The Bayesian analysis of the concatenated data sets was aligned over 877 bp (Figure 2). Although the analysis was run over 2,000,000 generations, the split frequencies never went below < 0.01. Nevertheless, posterior probabilities identified some very well supported clades. The ingroup was monophyletic (100%), and within it the first well-supported (100%) diverging lineage, called South II, that included specimens from SC, GA, and FL. The remaining samples clustered in a monophyletic lineage (93%) which did split into two sister clades, one (99%) including only southern ticks (South V) and one (100%) encompassing both southern and northern samples (American clade). Within South V, two lineages were identified. One of them, South I (100%), included samples from GA, FL, SC, NC, AL, and TX. The other one was split into two well supported branches (97%), South III (NC, SC, GA, and LA), and South IV (TX, LA). The American lineage, bifurcated basally into AM II (only southern specimens supported by 100% from AL, FL, LA, and SC) and the polytomic AM I which is represented by ticks collected in all of the 15 states. ML distance values, varied from 0% to 5.1% within clades, and from 1.6% (AMI vs. AMII) to 9.7% (South I vs. AMI).

Nuclear genes

A total of 156 ITS2 sequences (855 bp) were obtained corresponding to 144 unique genotypes. Sequences were very conserved (0.4-1.4% ML distance); mutations consisted of non-fixed singletons with no phylogenetic informative value. When cloned sequences obtained from individual ticks were examined, we realized that the two most divergent sequences (differing by 1.4% over an eight bp gap) were obtained from the same tick. ITS2 sequences were, therefore, not used for phylogenetic analysis. A total of 117 sequences (348 bp) of actin 1 were obtained (54 unique genotypes). Actin 1 sequences were even more conserved (0.00-0.035% ML distance) than ITS2 sequences and were not phylogenetically informative.

Genetic differentiation between mitochondrial clades

Only when missing data were transformed into null homozygotes, did the exact binomial test provide two significant tests out of 10 (*p*-value = 0.0861), which is not significant above the 5% expected under the null hypothesis. No pair of clades appeared significantly different with microsatellite markers after Bonferroni correction (smallest *p*-value = 0.1).

The Mantel test between matrices of mitochondrial and microsatellite distances (by individual ticks) did not yield a significant result ($r = 0.025$, *p*-value = 0.0964) and confirmed the results obtained by comparing genetic differences between clades. The absence of a clade signature on microsatellite loci allowed us to combine ticks from different mtDNA defined clusters for subsequent analyses.

Relevant levels of population subdivision

The effect of states within region was highly significant ($F_{SR} = 0.044$, *p*-value = 0.005) whereas the effect of region within the total sample was not significant ($F_{RT} = 0.017$, *p*-value = 0.86). Thus, the "state" subdivision, which in our case mostly corresponds to the "collection site" subdivision is the relevant unit to take into account.

Linkage disequilibrium, Hardy-Weinberg, and null alleles

Linkage disequilibrium tests detected nine out of 36 significantly linked pairs of loci (*p*-value = $5.349.10^{-5}$). Only five pairs remained significant after Bonferroni correction. There was a significant departure from Hardy-Weinberg expectations ($F_{IS} = 0.145$, *p*-value = 0.001). Figure 3 shows that this F_{IS} is strongly variable between loci, which led us to check for the presence of null alleles with MicroChecker. The MicroChecker analysis and subsequent exact binomial tests showed that null alleles can indeed explain the observed F_{is} values (all *p*-values > 0.0585).

Isolation by distance

There was a significant signature of isolation by distance in both C6 ($r = 0.582$ and *p*-value = 0.001) and C7 ($r = 0.704$ and *p*-value = 0.005) cohorts.

DISCUSSION

The initial controversy about *I. scapularis* potentially being two different species has mostly (Spielman et al. 1979; Oliver et al. 1993) been relegated to history. Nevertheless, the

molecular identification of an American clade (encompassing ticks from the whole distribution range of the tick), and of a "southern" clade, involving exclusively ticks found south of NC has maintained in the public and some of the scientific community the idea that northern and southern ticks are different, with northern ticks occupying the whole geographical area of *I. scapularis.* Nuclear analyses provided little clarification, as most nuclear genes proved to be very conserved and provided little phylogenetic information. Our mitochondrial data confirm the existence of several southern clades, not just one. One clade (South IV) is confined to the South-West (TX, LA), whereas the others (South I, II, and III) have generally a wider distribution. Unlike other studies (Norris et al. 1996; Qui et al. 2002; Humphrey et al. 2010; Sakamoto et al. 2014), the monophyly of the southern clades was not confirmed. Within the all-American clade, the basal split separates ticks from the 15 states we included in our sampling, from a lineage (AMII), which is only found in the South. The overall topology of the tree suggests a radiation of *I. scapularis* in the South, followed by dispersal towards the north of a single lineage as was previously reported (Norris et al. 1996; Qui et al. 2002; Humphrey et al. 2010; Sakamoto et al. 2014). The phylogeographical evolutionary history of this tick is the subject of a different study (Beati et al, unpublished).

Nevertheless, although nuclear gene analyses have generally indicated overall genetic homogeneity within *I. scapularis*, the results might depend on a number of factors, such as low mutation rates in the chosen genes and selective pressure constraining mutation rates. It was, therefore, necessary to compare mtDNA genetic distances with nuclear genetic distances obtained from molecular markers, microsatellite loci, known to be free from selective pressures, and characterized by high mutation rates (de Meeûs 2011). Our data do not reveal any correlation between mitochondrial and microsatellite distances, thus implying that mtDNA lineages have a different evolutionary history when compared to nuclear markers. This would confirm the previously reported discrepancies between mitochondrial and nuclear gene phylogenies (Norris et al. 1996; Mcclain et al. 1995; Sakamoto et al. 2014). Recent studies have revealed that high levels of mitochondrial diversity might merely reflect adaptive evolution (Bazin et al. 2006) and that mitochondrial gene evolution might not be neutral, as previously assumed (Gerber et al. 2001). Because of the lack of correlation between our mitochondrial and nuclear markers, in terms of population genetics, we can ignore the subdivision of our sample by mitochondrial clades. SNP analyses also revealed the occurrence of higher levels of genetic diversity in the

South, when compared with northern populations. Nevertheless, previsouly published loci (Van Zee et al. 2013) were mostly selected within housekeeping genes, which are also submitted to strong selective pressure and might not reflect neutral evolution. Moreover, SNPs are known to be frequently represented by bi-allelic loci only, a feature limiting their usefulness. They are also often plagued by heterogeneous mutation rates between alleles with a bias for transitions vs. transversions (Vignal et al. 2002).

Once we established that the genetic mtDNA subdivision did not correspond to real population subdivisions, we investigated the smallest level of subdivision we could detect in our data. Our results indicated that our samples were structured by state and not by region. As, once we selected the samples that could be used for this analysis, most of our states were represented by a single collection site, it would, therefore, be very important for future studies to collect from several sites within a single state and to establish real small-scale delimitations of subsamples. However, the results showed that the population units do not expand over state boundaries.

Linkage disequilibrium was present in our dataset, in particular between locus 16 and other loci, thus indicating that locus 16 might have to be eliminated from future studies. The deviations from Hardy-Weinberg revealed in several of our samples (Figure 3) could in part be explained by the occurrence of null alleles. By recoding null alleles as an additional non-null allele (999) with FreeNA, data could be corrected and used for the analysis of isolation by geographical distance.

It is evident, from our data, that genetic distances and geographical distances are correlated (Table 9 and 10). This is valid not only along a South-North latitudinal gradient, which was somewhat expected based on earlier controversies, but also along a longitudinal gradient with distances between FL and VA, being identical to distances between FL and TX.

CONCLUSION

In conclusion, mtDNA tree topologies can probably be regarded as present day evidence of population partitioning and local adaptation, long erased by secondary contact and re-merging that resulted in renewed gene flow. Within the distribution range of *I. scapularis*, subdivisions appear to be small and their real size can only be investigated by working at smaller geographical scales. The correlation between geographical distances and genetic distances in such a wide-

spread species, is not surprising. However, it also argues against the hypothesis of ticks being frequently carried up and down the Atlantic coast by birds, thus homogenizing those populations (Klich et al. 1996; Smith et al. 1996; Ogden et al. 2008). Although birds are well-known to transport ticks, even over very long distances, the ticks they move might not be able to survive once relocated. In addition, the fact that differences between south-western and south-eastern ticks are comparable to distances between northern and southern ticks implies that the genetic diversity follows a more complex pattern than a simple North-South split. This evidence mirrors the in-depth morphological analysis by Hutchenson et al. (1995) and Keirans et al. (1996), who demonstrated that there was not only a latitudinal, but also a longitudinal gradient of morphological diversity in *I. scapularis.*

Our data also strongly suggest that the reasons behind the discrepancy observed between the distributions of areas of high prevalence of Lyme disease and tick distribution are unlikely to be linked simply to tick genetic differences, but should rather find their explanation in environmental and ecological factors.

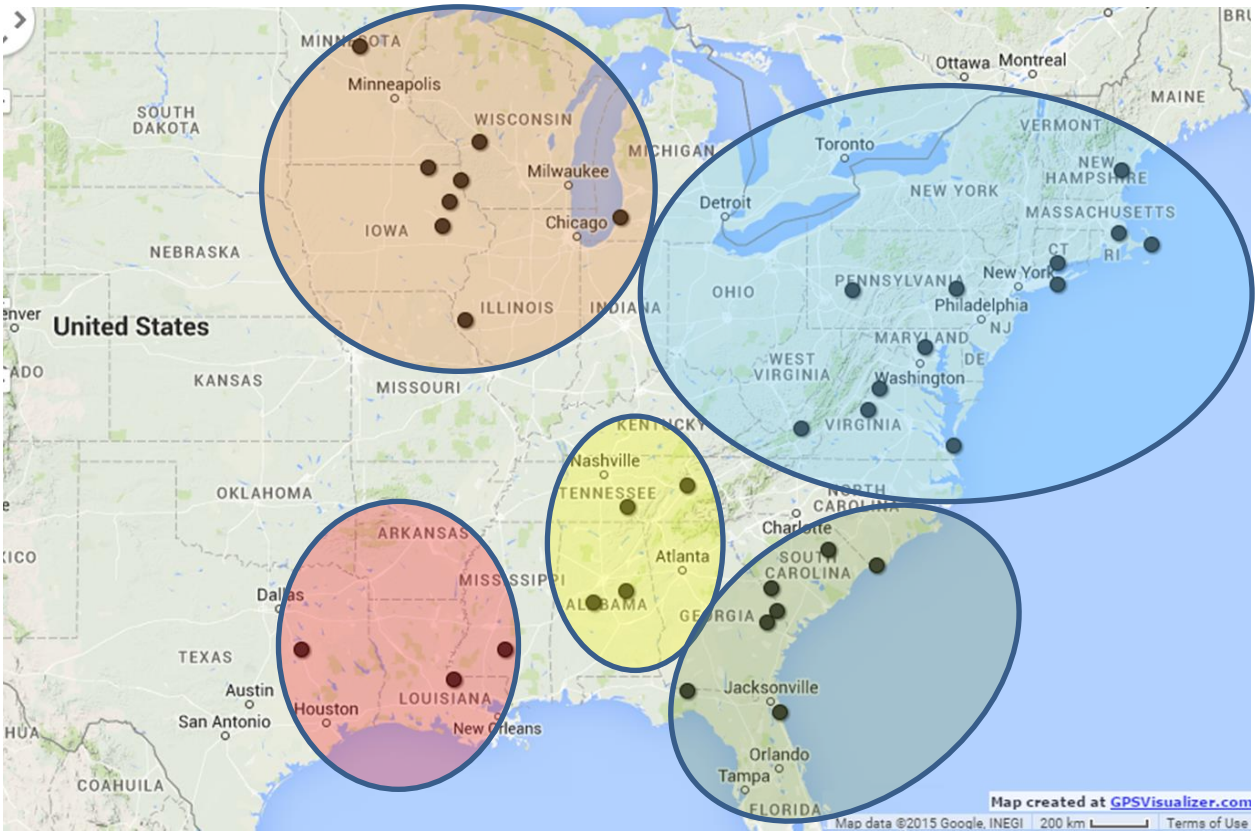

Figure 1. Map of collection sites. Information about geographic coordinates of each site and number of ticks collected at each site are listed in Table 1. Shaded circles represent regions used for testing population subdivision levels.

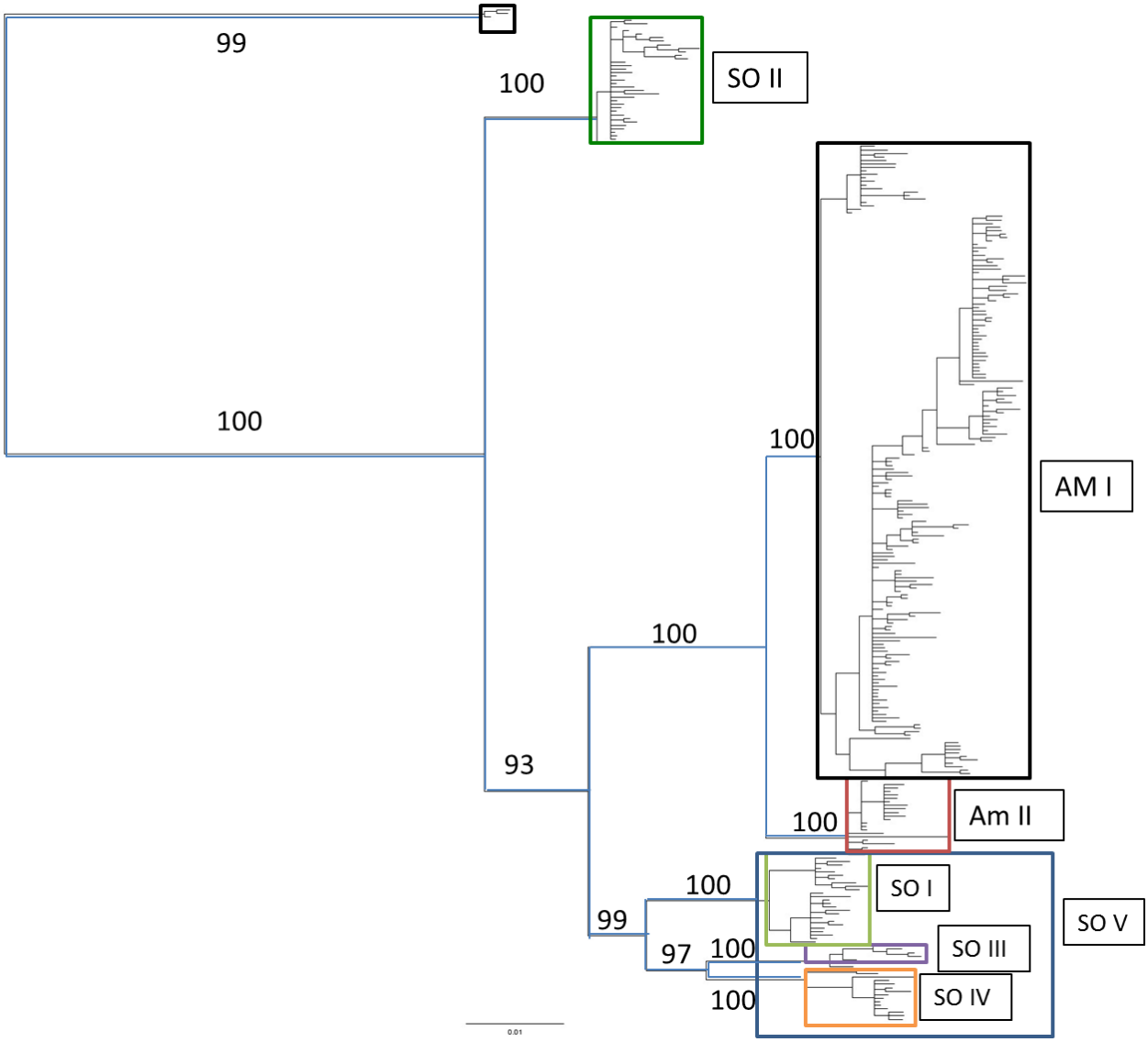

Figure 2. Bayesian tree of concatenates mtDNA gene sequences (12SrDNA and DL); posterior probabilities are shown over branches. Clade designation: AM I = American I, AM II = American II, SO I = South I, SO II = South II, SO III = South III, SO IV = South IV. The scale indicates the number of substitutions per site.

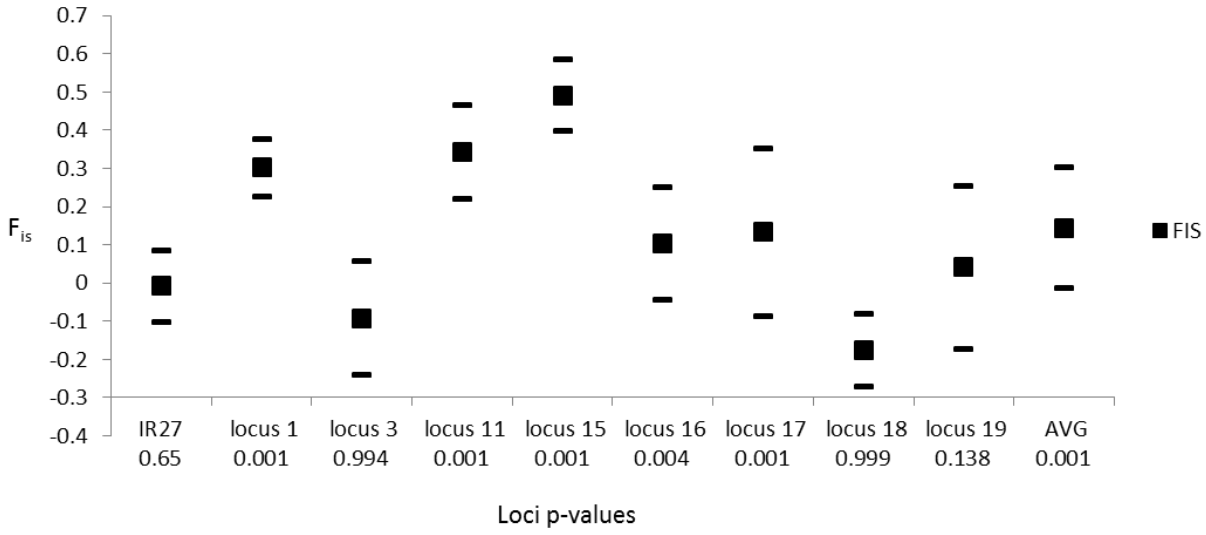

Figure 3. Graph of F_{is} values by locus with corresponding confidence intervals calculated by jackniving over standard errors. P-values were obtained by using Fstat and randomizing 1000 times alleles within samples.

Table 1. Geographic coordinates of collection sites and total number of collected ticks. Samples without cordinates not listed.

Region	Location/State	Latitude	Longitude	No. of ticks
North Coastal	Ice House/ Old Tree Site, CT	41.27	-72.81	31
	Cape Cod National Seashore, MA	41.71	-69.92	23
	Middleboro, MA	41.98	-70.91	1
	David W. Force Park, MD	39.29	-76.88	13
	William Floyd Estate, NY	40.76	-72.83	10
	Hawk Mountain Santuary, PA	40.67	-75.95	3
	Porcaro Estate, PA	40.64	-75.16	9
	First Landing State Park, VA	36.90	-76.01	1
	Shenandoah National Park, VA	38.29	78.29	16
	Mauricia Pet Grooming, VA	37.32	-80.73	4
	Oak Ridge Hunt Club, VA	37.79	-78.65	1
	Action Trading Post, ME	43.38	-70.85	1
North inland	Camp Ewahlu, IA	42.67	-91.59	3
	Pleasant Creek SRA, IA	42.11	-91.80	3
	Upper Iowa Accsss Wildlife Area, IA	43.46	-92.26	10
	Yellow River State Forest, IA	43.17	-91.25	1
	Van Buren State Park, MI	42.33	-86.30	2
	Fort McCoy, WI	44.03	-90.67	6
	Camp Ripley, MN	46.09	-94.34	1
	Liberty, IL	39.97	-91.1	1
South Coastal	Tall Timbers Research Station, FL	30.65	-84.22	10

	Guana River, FL	30.055	-81.37	7
	Beaver Pond, GA	32.73	-81.48	28
	Bulloch, GA	32.42	-81.79	1
	Shingletree Road Park, NC	33.91	-78.39	30
	Savannah River Site, SC	33.33	-81.66	58
	Darlington, SC	34.29	-79.87	5
South inland	Talladega National Forest, AL	33.24	-86.13	44
	Oakmulgee WMA, AL	32.95	-87.14	5
	Arnold Air Force Base, TN	35.37	-86.07	24
	Oak Ridge, TN	35.93	-84.26	1
	J & S Deer Processor, TX	31.72	-96.13	9
	Jefferson, MS	31.74	-89.89	1
	Weyanoke, LA	30.94	-91.46	8

Table 2. DNA samples used in this study, with geographic origin by state and life stages. Column "?" lists ticks that were too damaged during the sampling process to be correctly assigned to any life stage.

	Lifestage					
State	L	N	M	F	?	Total
AL	0	0	32	27	0	59
CT	0	7	13	12	0	32
FL	0	0	7	10	0	17
GA	0	1	10	25	0	36
IA	0	17	3	0	0	20
IL	0	1	0	0	0	1
LA	0	0	3	5	0	8
MA	11	7	3	3	3	27
MD	3	2	3	8	0	16
ME	0	1	0	0	0	1
MI	2	1	0	0	3	6
MN	0	1	0	0	0	1
MS	1	0	0	0	0	1
NC	0	0	8	22	0	30
NY	0	0	10	0	0	10
PA	3	7	19	19	0	48
SC	56	2	9	11	0	78
TN	0	1	8	17	0	26
TX	0	1	12	19	0	32
VA	10	1	5	7	0	23
WI	5	1	0	0	2	8
Total	91	51	145	185	8	480

Table 3. List of primers used in PCR for sequencing. Source of primer sequences: Beati et al. 2001 for 12SrDNA; Beati et al. 2012 for DL and ITS2; this study for actin.

Gene	Forward Primer 5' to 3'	Reverse Primer 5' to 3'
12SrDNA	T2A AATGAGAGCGACGGGCGATGT	T1B AAACTAGGATTAGATACCCT
Dloop	Dloop 3 TAACCGCTGCTGCTGGCACAA	Dloop 4 TAACCCTTTATTCAGGCAT
ITS2	Labru_B GTGAATTCTATGCTTAAATTCAGGGGGT	Zhnew TGAATTGCAGGACACACTGAGCAG
Actin 1	ACT5C_2F CGTACCACNGGTATCGTGCTC	ACT5C_2R GGATGCCGCACGATTCCATA

Table 4. List of primers used to amplify microsatellite loci and of the fluorescent dyes used to tag the forward primer. Source of primer sequences: Fagerberg et al. 2001 for L1; Delaye et al. 1998 for IR27; and Beati et al. (unpublished) for the remaining loci.

Locus	Forward Primer 5' to 3'	Reverse Primer 5' to 3'	Dye
L1	AAATGTCCGAACAGCCTTAT	GCCCTTGAGTCTACCCACTA	6FAM
L3	GCAGATCTCTTGGGCTAG	AAGCTAAGGCGTTCGTTG	VIC
L11	GCTTTAGCGGGCTGGT	TACGTGAATACGTCCTTGG	PET
L15	TATTGTAACCGACGCTAGG	GACAATCTCTACGCAAATCC	NED
L16	CCCCCAAACACGCACA	TTGCTTCATGCAGGGAAC	VIC
L17	CCAGCATTTAACCCTCAAG	TAGTGGGTATGGCACTG	HEX
L18	GTAGGTACCCTAAGAAGGAT	TTGAGGAAGCAGAATGTAGG	6FAM
L19	AGAACCAGTTCAGCATTCC	GAACATTTTCACGTGTTGC	PET
IR27	CATCGCTAGTGGCTAGAG	TTATAACCCGAGGTCGTAAAA	6FAM

Table 5. Composition of the PCR master mix used for amplifying DNA. As genotyping requires smaller volumes of amplicon when compared to sequencing, the total volume for genotyping reactions was reduced to 12.5 µl.

	Sequencing (25µl reaction) and genotyping (12.5 µl reaction)	
PCR master mix	dH_2O	10.3 µl
	Taq buffer (10X)	2.5 µl
	Taq master enhancer (5X)	5 µl
	$MgCl_2$ (25mM)	1.5 µl
	Primer 1(10pm/ µl)	1.25 µl
	Primer 2(10pm/ µl)	1.25 µl
	dNTPs (200mM each)	0.5 µl
	Taq DNA polymerase (5U/µl)	0.2 µl
	DNA sample	2.5 µl

Table 6. Thermocycler program used to amplify mitochondrial and nuclear genes.

	12S	Dloop	ITS2	M13	Actin
Program	94°C for 5’	94°C for 5’	94°C for 5’	94°C for 10’	94°C for 5’
Touchdown steps	5 cycles of 94°C for 20” 56°C-1°/cycle for 25” 70°C for 30”	8 cycles of 94°C for 20” 60°C-1°/cycle for 25” 72°-0.2°C/cycle for 40”	7 cycles of 94°C for 20” 65°C-2°/cycle for 20” 72°-0.2°C/cycle 20”		5 cycles of 94°C for 20” 56°C-1°/cycle for 20” 72.0°C°-0.5°C/cycle for 30”
	35 cycles of 94°C for 20” 50°C for 25” 70°C for 30”	30 cycles of 94°C for 20” 48°C for 45” 68°C for 40”	30 cycles of 94°C for 20” 50°C for 40” 70°C for 1’	30 cycles of 94°C for 45” 55°C for 20” 72°C for 30”	35 cycles of 94°C for 20” 55°C for 20” 72°C for 30”
Final elongation	70°C for 5”	68°C for 5”	70.0°C for 5’	72°C for 5’	72°C for 5’
Refrigeration and storage	04°C	04°C	04°C	04°C	04°C

Table 7. Thermocycler programs used to amplify microsatellite loci.

Loci	L11, L15	L1, L3, L16, L17, L18, L19	IR27
Program	94°C for 5’	93°C for 5’	94.0°C for 5’
Touchdown steps	5 cycles of 94°C for 20” 58°C-1.5°/cycle for 25” 72°C-0.4°/cycle for 30”	5 cycles of 93°C for 20” 55°C-1.5°/cycle for 30” 70°C-0.8°/cycle for 30”	5 cycles of 94°C for 20” 56°C-0.6°/cycle for 20” 72°C-0.6°/cycle for 30”
	30 cycles of 94°C for 20” 50°C for 30” 70°C for 30”	30 cycles of 93°C for 20” 47°C for 35” 67°C for 30”	30 cycles of 94°C for 20” 53°C for 30” 69°C for 30”
Final elongation	70°C for 5’	67°C for 5’	69°C for5’
Refrigeration and storage	04°C	04°C	04°C

Table 8. Genetic distances between main clades calculated form the mtDNA concatenated dataset with the maximum likelihood model.

	South IV	South III	South I	South II	AM II	AM I
South IV	0.00-0.0250					
South III	0.035-0.052	0.001-0.025				
South I	0.043-0.064	0.023-0.056	0.00-0.0310			
South II	0.060-0.078	0.063-0.085	0.036-0.074	0.000-0.290		
Am II	0.065-0.085	0.064-0.084	0.045-0.087	0.054-0.087	0.000-0.037	
AM I	0.049-0.087	0.060-0.096	0.032-0.097	0.042-0.086	0.016-0.069	0.000-0.051

Table 9. Cavalli-Sforza and Edwards genetic chord distances between states, obtained from cohort 6 microsatellite data.

	AL	FL	MD	PA	SC	TN
FL	0.41					
MD	0.51	0.60				
PA	0.58	0.63	0.60			
SC	0.41	0.45	0.57	0.50		
TN	0.35	0.46	0.51	0.55	0.44	
VA	0.47	0.55	0.51	0.50	0.50	0.41

Table 10. Cavalli-Sforza and Edwards genetic chord distances between states, obtained from cohort 7 microsatellite data.

	AL	FL	LA	SC	TX
FL	0.54				
LA	0.45	0.56			
SC	0.40	0.55	0.49		
TX	0.51	0.62	0.43	0.48	
VA	0.55	0.62	0.57	0.54	0.62

REFERENCES

Bazin E, Glemin S, Galtier N. 2006. Population size does not influence mitochondrial genetic diversity in animals. Science. 312: 570-71.

Beati L, Keirans JE. 2001. Analysis of the systematic relationships among ticks of the genera *Rhipicephalus* and *Boophilus* (Acari: Ixodidae) based on mitochondrial 12S ribosomal DNA gene sequences and morphological characters. J. Parasitol. 87(1):32-48.

Beati L, Patel J, Lucas-Williams H, Adakal H, Kanduma EG, Tembo-Mwase E, Krecek R, Mertins JW, Alfred JT, Kelly S, et al. 2012. Phylogeography and demographic history of *Amblyomma variegatum* (Fabricius) (Acari: Ixodidae), the tropical bont tick. Vector Borne Zoonotic Dis. 12(6):514-25.

Bonnet E, Van de Peer Y. 2002. Zt: a software tool for simple and partial mantel tests. J Stat Softw. 7(10):1-12.

Brookfield J. 1996. A simple new method for estimating null allele frequency from heterozygote deficiency. Mol Ecol. 5(3):453-5.

Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, Davis JP. 1982. Lyme disease-a tick-borne spirochetosis? Science. 216(4552):1317-9.

Cavalli-Sforza LL, Edwards AW. 1967. Phylogenetic analysis. Models and estimation procedures. Am J Hum Genet. 193(1):233-57.

Chapuis MP, Estoup A. 2007. Microsatellite null alleles and estimation of population differentiation. Mol Biol Evol. 24(3):621-31.

Clifford, CM, Anastos G, Elbl, A. 1961. The larval Ixodid ticks of the Eastern United States. Miscellaneous Publications of the Entomological Society of America. 2(3): 213-237.

Coombs JA, Letcher BH, Nislow KH. 2008. Create: a software to create input files from diploid genotypic data for 52 genetic software programs. Mol Ecol Resour. 8(3):578-80.

da Silva Vaz I Jr, Imamura S, Nakajima C, de Cardoso FC, Renard G, Masuda A, Ohasi K, Onuma M. 2005. Molecular cloning and sequence analysis of cDNAs encoding for *Boophilus microplus*, *Haemapahysalis longicornis* and *Rhipicephalus appendiculatus* actins. Vet Parasitol. 127(2):147-55.

de Meeûs T, Beati L, Delaye C, Aeschlimann A, Renaud F. 2002. Sex-biased genetic structure in the vector of Lyme disease, *Ixodes ricinus*. Evolution. 56(9):1802-1807.

de Meeûs T, Humair P, Grunau C, Delaye C, Renaud F. 2004. Non-mendelian transmission of alleles at microsatellite loci: an example in *Ixodes ricinus*, the vector of Lyme disease. Int J Parasitol. 34(8):943-50.

de Meeûs T. 2011. Initation A la genetique des population naturelles. Applications aux parasites et à leurs vecteurs. Marseilles France: IRD Éditions. 335p.

Delaye C, Aeschlimann A, Renaud F, Rosenthal B, de Meeûs T. 1998. Isolation and characterization of microsatellite markers in the *Ixodes ricinus* complex (Acari: Ixodidae). Mol Ecol. 7(3):360.

Durden L, Keirans JE. 1996. Nymphs of the genus *Ixodes* (Acari: Ixodidae) of the United States: taxonomy, identification key, distribution, hosts, and medical/ veterinary importance. Lanham, Maryland: Entomological Society of America. 95p.

Dieringer D, Schlötterer C. 2003. Microsatellite analyser (MSA): a platform independent analysis tool for large microsatellite data sets. Molecular Ecology Notes. 3(1):167-9.

Fagerberg AJ, Fulton RE, Black WC. 2001. Microsatellite loci are not abundant in all arthropod genomes: Analyses in the hard tick, *Ixodes scapularis* and the yellow fever mosquito, *Aedes aegypti*. Insect Mol Biol. 10(3):225-36.

Gerber A, Loggins R, Kumar S, Downling T. 2001. Does nonneutral evolution shape observed patterns of DNA variation in animal mitochondrial genomes? Annu Rev Genet. 35:539-66.

Goldstein DB, Schlötterer C. 1999. Microsatellites, evolution and application. Oxford, United Kingdom: Oxford University Press. 368p.

Goudet J. 2003. FSTAT (version 2.9.4), a program (for Windows 95 and above) to estimate and test population genetics parameters. (Available from the author upon request)

GOUDET J. 2005. Hierfstat, a package for r to compute and test hierarchical F-statistics. Mol Ecol Notes. 5(1):184-6.

GENEMARKER [computer program]. Version 2.4.0. State College, PA: SOFTGENETICS; 2012.

Holm S. 1979. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics. 6(2):65-70.

Hu LT. 2012. In the clicnic. Lyme disease. Ann Intern Med. 157(3):ITC2-2-ITC2-16.

Humphrey PT, Caporale DA, Brisson D. 2010. Uncoordinated phylogeography of *Borrelia burgdoferi* and its tick vector, *Ixodes scapularis*. Evolution. 64(9):2653-63.

Hutchenson HJ, Oliver JH, Houck MA, Strauss RE. 1995. Multivariate morphometric discrimination of nymphal and adult forms of the blacklegged tick (Acari: Ixodidae), a principal vector of the agent of Lyme disease in Eastern North America. J Med Entomol. 32(6):827-42.

Keirans JE, Hutcheson HJ, Durden LA, Klompen JS. 1996. *Ixodes* (*Ixodes*) *scapularis* (Acari:Ixodidae): redescription of all active stages, distribution, hosts, geographical variation, and medical and veterinary importance. J Med Entomol. 33(3):297-318.

Keirans JE, Clifford CM. 1978. The genus *Ixodes* in the United States: a scanning electron microscope study and key to the adults. J Med Entomol 15:1-38.

Kempf F, McCoy KD, de Meeûs T. 2010. Wahlund effects and sex-biased dispersal in *Ixodes ricinus*, the European vector of Lyme *Borreliosis*: New tools for old data. Infect Genet Evol. 10(7):989-97.

Klich M, Lankester MW, Wu KW. 1996. Spring migratory birds (Aves) extend the northern occurrence of blacklegged tick (Acari: Ixodidae). J Med Entomol. 33(4): 581-5.

Lehmann T, Hawley WA, Kamau L, Fontenille D, Simard F, Collins FH. 1996. Genetic differentiation of *Anopheles gambiae* populations from East and West Africa: comparison of microsatellites and allozyme loci. Heredity. 77(2):192-200.

Librado P, Rozas J. 2009. DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. Bioinformatics. 25(11):1451-52.

Maddison DR, Maddison WP. MACCLADE [computer program]. Version 4.08. Sunderland, Massachusetts: Sinauer Associates; 2005. MacClade 4: Analysis of phylogeny and character evolution.

McLain DK, Wesson DM, Oliver JH Jr, Collins FH. 1995. Variation in ribosomal DNA internal transcribed spacers 1 among eastern populations of *Ixodes scapularis* (acari: Ixodidae). J Med Entomol. 32(3)353-60.

Meltzer MI, Dennis DT, Orloski KA. 1999. The cost effectiveness of vaccinating against Lyme disease. Emerg Infect Dis. 5(3):321-8.

Morgan E. 2008. Genepop on the web. Available at http://genepop.curtin.edu.au/.

Norris DE, Black WC, Keirans JE, Klompen JSH. 1996. Population genetics of *Ixodes scapularis* (Acari: Ixodidae) based on mitochondrial 16S and 12S genes. J Med Entomol. 33(1):78-89.

Ogden NH, Lindsay LR, Hanincova K, Barker IK, Bigras-Poulin M, Charron DF, Heagy A, Francis CM, O'Callaghan CJ, Schwartz I, Thompson RA. 2008. Role of migratory birds in

introduction and range expansion of *Ixodes scapularis* ticks of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. Appl Environ Microbiol. 74(6):1780-1790.

Oliver JH Jr, Owsley MR, Hutcheson HJ, James AM, Chen C, Irby WS, Dotson EM, McLain DK. 1993. Conspecificity of the ticks *Ixodes scapularis* and *I. dammini* (Acari: Ixodidae). J Med Entomol. 30(1):54-63.

Qiu WG, Dykhuizen DE, Acosta MS, Luft BJ. 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(3):833-49.

R [computer program]. Version 3.2.0. Vienna, Austria: R Foundation for Statistical Computing; 2015. R: a language and environment for statistical computing.

Raymond M, Rousset F. 1995. Genepop (version 1.2): population genetics software for exact test and ecumenicism. J Heredity. 86:248-249.

Rich SM, Rosenthal BM, Telford SR, Spielman A, Harti DL, Ayala FJ. 1997. Heterogeneity of the internal transcribed spacer (ITS-2) region within individual deer ticks. Insect Mol Biol. 6(2):123-9.

Ronquist F, Huelsenbeck JP. 2003. MRBAYES 3: bayesian phylogenetic inference under mixed models. Bioinformatics 19(12):1572-1574.

Rosenthal BM, Spielman A. 2004. Reduced variation among northern deer tick populations at an autosomal microsatellite locus. J Vector Ecol. 29(2):227-35.

Rousset F. 2008. Genepop'007: a complete reimplementation of the Genepop software for Windows and Linux. Mol Ecol Resources. 8(1): 103-6.

Sakamoto JM, Goddard J, Rasgon JL. 2014. Population and demographic structure of *Ixodes scapularis* say in the eastern United States. PLoS One. 9(7):e101389.

SEQUENCHER [computer program]. version 5.4. Ann Arbor, Michigan: Gene Codes Coporation; 2013. sequence analysis software.

Smith RP, Rand PW, Lacombe EH, Morris SR, Holmes DW, Caporale DA. 1996. Role of bird migration in the long-distance dispersal of *Ixodes dammini*, the vector of Lyme disease. J Infect Dis. 174(1):221-224.

Spielman A, Clifford CM, Piesman J, Corwin MD. 1979. Human babesiosis on Nantucket island, USA: description of the vector, *Ixodes (Ixodes) dammini*, n. sp. (Acarina: Ixodidae). J Med Entomol. 15(3):218-34.

Stanek G, Wormser GP, Gray J, Strle F. 2012. Lyme borreliosis. Lancet. 379(9814):461-73.

Steer AC, Coburn J, Glickstein L. 2004. The emergence of Lyme disease. J Clin Invest. 113(8):1093-101.

Swofford DL. PAUP* [computer program]. Version 4 Beta 10. Sunderland, Massachusetts: Sinauer Associates; 2015. Phylogenetic Analysis Using Parsimony (*and Other Methods).

Van Oosterhout C, Hutchinson WF, Wills DP, Shipley P. 2004. MICRO-CHECKER: software for identifying and correcting genotyping errors in microsatellite data. Molecular Ecology Notes. 4(3):535-8.

Van Zee JP, Geraci NS, Guerrero FD, Wikel SK, Stuart JJ, Nene VM, Hill CA. 2007. Tick genomics: the *Ixodes* genome project and beyond. Int J Parasitol. 37(12):1297-305.

Van Zee J, Black WC, Levin M, Goddard J, Smith J, Piesman J. 2013. High SNP density in the blacklegged tick, *Ixodes scapularis*, the principal vector of Lyme disease spirochetes. Ticks Tick Borne Dis. 4(1-2):63-71.

Vignal A, Milan D, SanCristobal M, Eggen A. 2002. A review on SNP and other types of molecular markers and their use in animal genetics. Genet Sel Evol. 34(3):275-305.

Vink CJ, Hedin M, Bodner MR, Maddison WP, Hayashi CY, Garb JE. 2008. Actin 5C, a promising nuclear gene for spider phylogenetics. Mol Phylogenet Evol. 48(1):377-82.

Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38(6):1358-70.

Wesson DM, McLain DK, Oliver JH, Piesman J, Collins FH. 1993. Investigation of the validity of species status of *Ixodes dammini* (Acari: Ixodidae) using rDNA. Proc Natl Acad Sci U S A. 90(21):10221-5.

Wright S. 1965. The interpretation of population structure by F-statistics with special regard to systems of mating. Evolution 19(3):395-420.

Xu G, Fang QQ, Keirans JE, Durden LA. 2003. Molecular phylogenetic analyses indicates that the *Ixodes ricinus* complex is a paraphyletic group. J Parasitol. 89(3):452-7.

Zhang DX, Hewitt G.1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol. 25(2): 99-120.

Zhang X, Meltzer MI, César AP, Hopkins AB, Wroth L, Fix AD. 2006. Economic impact of Lyme disease. Emerg Infect Dis. 12(4):653-60.