Université de Montréal

Utilisation des outils phylogéographiques pour explorer la diversité génétique de *Borrelia burgdorferi* et le paysage génétique de la maladie de Lyme au Canada

Par SAMIR MECHAI

Département de pathologie et microbiologie
Faculté de médecine vétérinaire

Thèse présentée à la Faculté de médecine vétérinaire
en vue de l'obtention du grade de *Philosophiae Doctor* (Ph. D.)
en sciences vétérinaires option épidémiologie

Avril 2017

Résumé

La maladie de Lyme est la maladie infectieuse zoonotique la plus rapportée en zone tempérée. Elle a des conséquences importantes sur la santé humaine. Au Canada, elle est considérée comme étant émergente du fait principalement de l'expansion de la tique vectrice *Ixodes scapularis* vers de nouvelles zones du Sud Canadien où elle s'est établie. Par conséquent, les aires endémiques de la maladie de Lyme ne cessent de s'élargir, favorisées notamment par les changements climatiques. Ces événements coïncident avec la radiation adaptative que montre actuellement la bactérie *Borrelia burgdorferi (B. burgdorferi)*, qui est responsable des cas de maladie de Lyme en Amérique du Nord.

Devant une telle problématique, les structures phylogéographiques de l'agent pathogène et de son vecteur montrent étonnamment peu de corrélation, malgré leur lien écologique intime. L'évaluation du risque environnemental de la maladie de Lyme nécessite d'avoir une image claire des différents processus qui œuvrent et qui guident l'expansion du pathogène et de son vecteur. Cependant cette image devient beaucoup plus complexe si on considère la diversité génétique de *B. burgdorferi*. En effet, elle peut avoir des effets négatifs sur les tests de diagnostics actuels et du fait de différents niveaux de pathogénicité, rendre le tableau clinique complexe.

L'une des réponses à ces problématiques peut se trouver dans la génétique bactérienne de ce pathogène et sa relation avec son vecteur et ses hôtes. Dans cette étude, nous investiguons la diversité et la structure génétique des populations de *B. burgdorferi* sensu stricto dans le Sud du Canada et dans le Nord des États-Unis (ÉU) par l'utilisation des outils phylogéographiques, à savoir : i) la phylogénie pour explorer l'histoire évolutive de *B. burgdorferi*, et ii) la modélisation géo-spatiale pour définir sa structure génétique actuelle à l'échelle spatiale.

Au total, 750 échantillons de *B. burgdorferi* sont exploités pour la reconstruction des liens de parenté entre les différentes souches typées avec la méthode multi-locus du typage des séquences (MLST) (477 échantillons des ÉU et du Canada sont disponibles dans la base *pubmlst.org/bburgdorferi/* et 273 échantillons canadiens, utilisés pour la première fois, proviennent de la surveillance active). Deux autres approches complémentaires sont utilisées

dans cette étude pour caractériser les différentes souches, à savoir le gène du plasmide cp26 codant pour la protéine C de surface (*ospC*) et l'espace inter-génique ribosomal *rrs-rrlA* (IGS).

Nos travaux ont montré que la diversité génétique de *B. burgdorferi* au Canada est relativement semblable à celle des ÉU, mais que seulement 1/5 des souches sont communes entre les deux pays. L'image phylogéographique de *B. burgdorferi* en Amérique du Nord apparait plus complexe que ce qui était connu jusqu'à présent, tant sur le plan géographique que génétique. En effet, des souches de diverses origines géographiques forment des clades significatifs alors que d'autres qui ont les mêmes origines géographiques sont éparpillées sur différents clades (génétiquement très distinctes). Ceci est un signal probable de populations refuges. De même, une complexité génétique a été enregistrée, puisque 90 % des échanges génétiques se font à l'intérieur de chaque sous-structure phylogénétique.

Une telle structure génétique peut être l'œuvre d'un patron écologique. Les fortes associations des génotypes avec des hôtes spécifiques peuvent limiter le flux génétique entre ces sous-structures. Dans ce contexte, la modélisation statistique a permis d'identifier des associations entre génotypes et certaines espèces de rongeurs : CC34 et ospC G avec le tamia rayé ; CC403, ospC A et RST1 avec la souris à pattes-blanches ; et CC4, ospC H et IGS 2D avec la souris sylvestre. La phylogénie a montré que les souches associées aux tamias sont plus anciennes comparativement à celles associées aux souris qui semblent avoir évoluées plus récemment. Des études basées sur le fossile du tamia corroborent cette hypothèse. En effet, cette espèce était l'une des rares espèces de petits mammifères ayant survécu durant les périodes glaciaires-interglaciaires. Durant son histoire, elle a entrepris une migration Nord-sud, alors que les souris, plus fragiles aux conditions climatiques ont entrepris une migration Sud-nord suite au retrait de la couche glaciaire.

Les adaptations des souches de *B. burgdorferi* aux hôtes peuvent avoir été le moteur de différences de pathogénicité des génotypes en Amérique du Nord. Elles façonnent aussi la phylogéographie contemporaine de ce pathogène dans cette partie du monde. En effet, la modélisation géo-spatiale à l'aide des probabilités conditionnelles a révélé que des souches (p.ex. ST1) présentent des modèles d'occurrence déterminés en partie par la connectivité des forêts.

Cette étude offre une première image phylogéographique compréhensive de *B. burgdorferi* dans le Sud du Canada, nécessaire pour comprendre l'épidémiologie évolutive de

la maladie de Lyme en Amérique du Nord. Elle montre qu'il existe une diversité génétique importante, ce qui peut aider à comprendre les variations géographiques des souches pathogènes et les méthodes de dispersion de *B. burgdorferi* en Amérique du Nord.

Ces résultats permettront d'améliorer notre habilité à prédire le risque de la maladie de Lyme au Canada et aideront à développer de nouvelles méthodes de diagnostic.

Mots-clés : *Borrelia burgdorferi*, maladie de Lyme, phylogéographie, histoire évolutive, typage multi-locus, connectivité du paysage, modélisation spatiale.

Abstract

Lyme borreliosis is the most reported zoonotic infectious disease in temperate zones with significant consequences for human health. In Canada, it is considered to be emerging due to the expansion of the *Ixodes scapularis* tick vector to new areas in southern Canada where it has become established. Therefore, endemic areas of Lyme disease continue to grow, and are being promoted by climate change. These events coincide with adaptive radiation of *B. burgdorferi* in North America.

Within this problem, the phylogeographic pictures of the pathogen and its vector show surprisingly little correlation despite their intimate ecological association. The environmental risk assessment of Lyme disease requires a clear comprehension of the different patterns acting and driving the expansion of the pathogen and its vector. This picture is made more complex by considering the genetic diversity of *B. burgdorferi* which is important to its probable effects on diagnostic test performance and the differential pathogenicity amongst strains.

Some answers to these problems may be found by study of the bacterial genetics of this pathogen to improve knowledge of its relationship with its vector and hosts. In this study we investigated the diversity and the genetic structure of *B. burgdorferi* sensu stricto populations in southern Canada and northern US by using phylogeographic tools: i) phylogenetic methods to explore the historical and the evolutionary processes of *B. burgdorferi* and ii) geo-spatial modeling to define the spatial distribution of different strains and their patterns of spread.

A total of 750 samples of *B. burgdorferi* (477 samples are from US and Canada which are available in the *pubmlst.org/bburgdorferi/*, and 273 samples collected using active surveillance in Canada and analyzed for the first time in this study) were used to reconstruct the phylogenetic relationship between different strains genotyped using MLST approach. Two other complementary approaches were used to strain-type *B. burgdorferi*, the outer surface protein C (*ospC*) and the ribosomal spacer *rrs-rrlA* (IGS).

It was found that the genetic diversity of *B. burgdorferi* in Canada is relatively similar to that of the US but only 1/5 strains are common to the both countries. This shows that the phylogeographic image of *B. burgdorferi* in North America appears to be more complex than

previously known: i) geographic complexity because strains of various geographical origins form significant clades and others which have the same geographical origin but are genetically more distant dispersed on different clades (this is a probable signal of refugial populations) and ii) a genetic complexity since 90% of genetic exchanges occur within closely related phylogenetic groups. Such genetic structure can be driven by an ecological process, and we hypothesized that the associations of *B. burgdorferi* genotypes with specific hosts reduced gene flow between different groups. Statistical modeling allowed the identification of different associations between genotypes and certain rodent host species: CC34 and *ospC* G with Eastern Chipmunk; CC403, *ospC* A and RST1 with the white-footed mice; and CC4, *ospC* H and IGS 2D with the deer mice.

However, the phylogeny also showed that strains associated with chipmunk are older relatively to those associated with mice, which appear to have evolved more recently.

Studies based on the chipmunk fossil data in North America corroborate this hypothesis. In fact, this species was one of the few small mammals that survived during the glacial-interglacial periods which thought allowed the *B. burgdorferi* surviving. It had undergone a southward migration, whereas mice known to be more sensitive to climatic conditions had experienced a Northward migration after the last Pleistocene which led all these small mammals to share a common history.

Therefore, host adaptations may have shaped the contemporary phylogeography of the pathogen in North America, which may be the driving force behind the differential pathogenicity of *B. burgdorferi* genotypes.

Geo-spatial modeling using conditional probabilities showed that rodent-associated strains (e.g. ST1 of CC403 that is associated with white-footed mice) exhibit specific patterns of occurrence of dispersal driven by the landscape characteristics of forest connectivity.

This study provides i) a first comprehensive phylogeographic picture of *B. burgdorferi* in southern Canada that is essential to understand the evolutionary epidemiology of Lyme disease in Northern America, ii) knowledge of the high genetic diversity in Canada that may be key to understanding geographic variations in occurrence of pathogenic strains and improving the specificity and sensitivity of diagnostic tests, and iii) understanding the methods of dispersion of *B. burgdorferi* in Northern America. Together these findings will improve our

ability to predict risk from Lyme disease in Canada, and will drive research into new diagnostic methods.

Keywords: *Borrelia burgdorferi*, Lyme disease, phylogeography, evolutionary history, multi-locus typing, landscape connectivity, spatial modeling.

Table des matières

Liste des tableaux

Liste des figures

Liste des sigles

°C: Degré Celsius

AIC: Akaike's information criterion

CDC: Centers for Disease Control and Prevention

CSD: Census subdivision

DNA: Deoxyribonucleic acid

ELISA: Enzyme-Linked Immunosorbent Assay

GLM: Generalized linear model

GPS: Global positioning system

IGS: Intergenic spacer

MLST: Multi-locus sequence typing

ospC: Outer surface protein C

ospA: Outer surface protein A

ospB: Outer surface protein B

LR: Likelihood ratio

LB: Lyme borreliosis

NML: National microbiology laboratory

PC: probability of connectivity

PCR: Polymerase chain reaction

RST: Ribosomal sequence type

SE: Standard error

SIG-GIS : Système d'information géographique – Geographic information system

TROSPA : Récepteur de la tique à la protéine de surface A

USA: United States of America

ÉU: États-Unis

VIH: virus de l'immunodéficience humaine

p.ex. : par exemple

e.g. : *exemplī grātia* ou par exemple

i.e. : *id est* ou c'est-à-dire

À mes parents et à toute ma famille

Remerciements

J’aimerais en premier lieu adresser mes remerciements les plus vifs à mon directeur de recherche, le Dr. Nicholas Ogden, pour l’intéressant sujet de doctorat qu’il m’a proposé. Merci de m’avoir soutenu moralement et financièrement, orienté, guidé sans cesse tout au long de ce projet. Chaque geste, chaque mot de sa part avait un sens qui me permettait d’apprendre à chaque instant et d’aller de l’avant. Il a été divin et sa générosité, sa confiance, son optimisme étaient tout ce qu’il fallait pour que ce projet aboutisse. Merci.

Merci également à mon co-directeur de recherche, le Dr. Pascal Michel, pour ses encouragements, ses conseils précieux et son encadrement. Merci.

Je tiens aussi à envoyer mes précieux remerciements à nos collaborateurs, Gabrielle Margos, Edward Feil, Robbin Lindsay et son équipe du NML pour leurs efforts et leurs précieuses contributions qui ont permis à ce travail de se concrétiser. Grand merci.

Je remercie aussi mon comité-conseil pour leurs orientations et conseils aussi précieux qu’importants pour la construction de ce projet. Merci.

Également merci au Dr. Serge-Olivier Kotchi et à Yann Pelcat pour leurs précieuses aides en géomatique, au Dr. Audrey Simon pour avoir accepté de lire ce document ainsi qu’à l’équipe du GREZOSP pour le partage.

Merci à toute ma famille, sans votre soutien je n’y serais jamais arrivé.

Enfin, je tiens à remercier vivement les membres du jury : le Dr. Patrick leighton pour avoir accepté de présider le jury, le Dr. John Morris Fairbrother et le Dr. Jean Tsao pour avoir accepté de faire partie du jury de cette thèse et d’examiner ce travail.

I. Introduction

Après avoir prédit la fin de l'ère des maladies infectieuses vers la fin des années 1970, certains experts ont suggéré de se concentrer sur les maladies chroniques telles que le cancer et les maladies cardiaques (World Health Organization, 2000). Cependant, une trentaine d'années plus tard, le combat contre les maladies infectieuses en général et celles émergentes en particulier reste encore d'actualité. Il est considéré comme étant l'un des plus grands défis de l'humanité à l'aube de l'ère de la mondialisation (Fauci, 2005; Harris et Reza, 2012). Les maladies infectieuses sont responsables de 63 % des décès infantiles à travers le monde (Alarcón et al., 2010). Environ 60 % d'entre elles sont des zoonoses dont 72 % proviennent des animaux sauvages (Alarcón et al., 2010; World Health Organization, 2000).

L'Organisation Mondiale de la Santé (OMS) les classe en 4 grandes catégories : i) maladie due à un agent infectieux connu mais nouvellement adapté à l'homme (p.ex. Ébola), ii) maladie tout à fait nouvelle (p.ex. syndrome respiratoire aigu sévère «SARS»), iii) maladie réémergente (p.ex. Malaria, Tuberculose, Dengue, Chikungunya, fièvre West Nile, maladie de Lyme), iv) maladie causée par un agent infectieux ayant changé son mode de transmission (p.ex. Chagas, Nipah) (Harris et Reza, 2012).

L'émergence et la réémergence d'une maladie ainsi que sa dissémination au sein d'une population ou dans une région donnée est souvent le résultat d'une combinaison de plusieurs facteurs biotiques et abiotiques liés à l'environnement immédiat de l'agent infectieux, de son vecteur et/ou de son hôte (Jones et al., 2008). Cet environnement est directement influencé par les changements globaux que subit le monde actuel (p.ex. réchauffement climatique). Dans ce contexte, l'OMS et l'Organisation mondiale de la santé animale (OIE) ont dû repenser le système de surveillance des maladies infectieuses, d'où l'avènement du paradigme «Une Seule Santé ou One-Heath » (Zinsstag et al., 2011).

La borréliose de Lyme (BL), communément appelée la maladie de Lyme est l'un des meilleurs exemples avancés par les défenseurs de cette approche pour illustrer l'importance de considérer ce paradigme (Fialkowski et al., 2016). La maladie de Lyme est une maladie infectieuse zoonotique transmise à l'hôte à la suite d'une morsure par une tique dure infectée du genre *Ixodes* (Houwerzyl et al., 1984; Matuschka et Spielman, 1986; Wood et Lafferty, 2013). Suite à l'infection, différents hôtes mammifères développent la maladie dont l'humain

et certains animaux domestiques tels que le chien et le cheval (Appel, 1990; Divers et al., 2001; de la Fuente et al., 2008).

Alors que cette maladie est connue depuis le début des années 1900 (Houwerzyl et al., 1984; Matuschka et Spielman, 1986; Wood et Lafferty, 2013), elle n'a été décrite d'une façon plus précise qu'au milieu des années 1970 suite à l'épidémie qui a frappé la région du Lyme au Connecticut dans le Nord-Est des États-Unis (Kurtenbach et al., 2006). Brièvement, au stade primaire, la maladie se manifeste souvent par une fièvre, des maux de tête et de fatigue similaires aux symptômes du syndrome grippal (Smith et al., 2002). La maladie est surtout diagnostiquée dans 70 à 80 % des cas grâce à une lésion cutanée qui apparaît autour de la morsure de la tique dite érythème migrant (EM) (Steere et al., 2004). En l'absence de traitement, la maladie peut évoluer vers des stades secondaires avec des manifestations neurologiques (p.ex. méningoradiculite, paralysie faciale), articulaires et plus rarement cardiaques (p.ex. troubles de conduction) (Van der Linde, 1991).

L'agent causal de la BL est une bactérie de la famille des spirochètes, *Borrelia burgdorferi,* isolée pour la première fois par Willy Burgdorfer en 1981 (Burgdorfer et al., 1982). À la suite d'analyses génétiques approfondies, il s'est avéré que la BL est associée à un complexe d'espèces appelé le complexe de *B. burgdorferi* sensu lato, qui comprend une vingtaine espèces (Steere et al., 2004; Postic et al., 2007; Rudenko et al., 2009; Becker et al., 2016). Cinq d'entre elles sont pathogènes pour l'humain : *B. burgdorferi* sensu stricto (ss), *B. afzelii*, *B. garinii*, *B. spielmanii* et *B. bavariensis* (Postic et al., 2007). Alors que six autres sont potentiellement pathogènes : *B. bissettii*, *B. kurtenbachii*, *B. lusitaniae*, *B. valaisiana* et *B. miyamotoi* (Rudenko et al., 2011; Krause et al., 2013).

La BL touche l'hémisphère Nord (Amérique du Nord, Europe, Asie et l'Afrique du Nord) (Kurtenbach et al., 2006), mais elle est régulièrement associée à une seule espèce en Amérique du Nord, à savoir *B. burgdorferi* ss (citée *B. burgdorferi* tout au long de ce document) (Margos et al., 2010). La transmission de cette bactérie est assurée par la tique à pattes-noires, *Ixodes scapularis* dans le Nord-est et le Midwest (incluant le Sud du Canada allant du Manitoba aux Maritimes), mais aussi par la tique à pattes-noires occidentale *Ixodes pacificus* sur les côtes ouest des États-Unis (ÉU) et en Colombie Britannique (Ogden et al., 2009; Humphrey et al., 2010; Margos et al., 2012).

En Amérique du Nord, les ÉU ont classé cette maladie parmi les zoonoses non alimentaires prioritaires en termes de surveillance du fait de son caractère émergent et de sa gravité (Qiu et al., 2002). Elle a atteint des incidences allant jusqu'à 300,000 nouveaux cas par ans, dont 80% sont enregistrés dans le Nord-est (Mead, 2015). Au Canada, le nombre de cas humains importés reste relativement stable depuis 1994, quand la plupart des cas étaient acquis hors du pays. Selon l'agence de la santé publique du Canada (ASPC), les cas endémiques par contre sont en net augmentation, passant de 40 cas en 2004 à 917 cas en 2015 (http://healthycanadians.gc.ca/).

Dans le contexte alarmant de maladie de Lyme que connaissent les ÉU, qu'en est-il de la situation épidémiologique actuelle de ce pathogène au Canada?

La BL reste une maladie émergente au Canada, mais aux conséquences importantes pour la santé humaine. L'agent pathogène arrive principalement des régions du Midwest et du Nord-est des ÉU grâce aux tiques *I. scapularis* transportées en partie par des oiseaux migrateurs, comme montré par des études sur les oiseaux (Ogden et al., 2008a) et les données de la surveillance passive (Ogden et al., 2009). Ces données ont montré également que les souches dans les tiques qui arrivent au Canada sont relativement les mêmes souches identifiées aux ÉU (Ogden et al., 2011).

Dans ce contexte, en se basant sur les données de la surveillance active menée par l'ASPC depuis plus 10, quelle est l'image phylogéographique de cette bactérie? Y a-t-il un patron clair qui expliquerait son expansion sur le territoire canadien?

De plus en plus d'études spéculent sur le rôle des hôtes terrestres dans la dispersion de *B. burgdorferi* (Ogden et al., 2013a; Estrada-Peña & de la Fuente, 2014; Estrada-Peña et al., 2016). En effet, si le mode de dispersion de la tique est assez bien connu, le mécanisme d'expansion de la bactérie *B. burgdorferi* à travers le Canada reste très mal connu. Est-ce que cette expansion est mieux expliquée par la dispersion des tiques dont la mobilité est assurée par les hôtes (p.ex. les oiseaux) ? Ou bien par les hôtes eux même du fait que certains d'entre eux sont des réservoirs de l'agent pathogène?

B. burgdorferi est un endoparasite obligatoire qui vit et se transmet grâce à son vecteur et sa gamme d'hôtes. La définition du rôle de chacune de ces composantes est d'un intérêt capital pour la compréhension de l'épidémiologie de la maladie de Lyme dont dépend le système de surveillance de cette maladie au Canada.

Globalement la problématique de la maladie de Lyme est complexe et polymorphe tant sur le plan de la recherche scientifique que sur les plans clinique et médical, et sur les plans de la mise en place de politiques de la santé publique.

La faible sensibilité que montre les tests de laboratoire disponibles durant le stade précoce de l'infection (Aguero-Rosenfeld et al., 2005; Allen et al., 2016) et la persistance des symptômes dans certains cas après une antibiothérapie (Koedel et Pfister, 2017) compliquent davantage l'établissement d'un diagnostic clair et efficace.

Il y a de plus en plus un besoin pressant en termes de politiques de santé publique pour évaluer les risques environnementaux que représente la maladie de Lyme. D'autant plus que récemment, le Centre pour le contrôle et la prévention des maladies (CDC) l'a classée comme étant la maladie à développement le plus rapide après le virus de l'immunodéficience humaine (VIH) aux ÉU, avec une incidence 10 fois plus élevée que ce qui était connu jusque-là (Kuehn, 2013).

L'élaboration d'un système de surveillance sensible et efficace à l'échelle fédérale est l'une des lignes directrices de la loi Lyme (L.C. 2014, ch. 37) au Canada. Dernièrement, les États-Unis et la France se sont engagés eux aussi sérieusement à élaborer des plans de surveillance de cette maladie (respectivement la Loi sur les cures du 21ème siècle aux ÉU et le Plan national de lutte contre la maladie de Lyme et les maladies transmises par les tiques en France), plus axés sur la prise en charge du patient, la compréhension de cette maladie et la prévention.

La mise en place de ces plans se base sur la capacité de chaque pays à identifier et à classer les cas de Lyme, et à produire les connaissances à transmettre aux praticiens de la santé et aux administrations en charges de la prévention. Une évaluation du risque adéquate est capitale pour mieux protéger efficacement les populations humaines. Ce qui nécessite de pouvoir identifier les cas de façon précoce. Selon le CDC, la mauvaise classification des cas humains est un biais fréquent des systèmes de surveillance, c'est-à-dire que la sous-déclaration des cas humains est plus susceptible de se produire dans des zones fortement endémiques, alors que la sur-déclaration est plus susceptible de se produire dans des zones non-endémiques (Bacon et al., 2008; Tuite et al., 2013).

Par ailleurs, l'évaluation du risque ne doit pas se baser uniquement sur la présence du vecteur et sur la déclaration des cas humains, ce qui pourrait fausser l'évaluation. En effet, on sait maintenant grâce aux études phylogéographiques que les occurrences du vecteur *I. scapularis* et de l'agent pathogène *B. burgdorferi* sont faiblement corrélées en Amérique du Nord (p.ex. le NE versus le Sud des ÉU) (Humphrey et al., 2010; Ogden et al., 2011).

La compréhension de l'épidémiologie de la BL fournira de l'information pour alimenter l'algorithme du système de surveillance de sorte que : i) l'identification des souches de *B. burgdorferi* dans l'environnement permettra d'identifier le niveau et le type du danger, ii) la définition des relations entre les différentes souches permettra de comprendre comment ce pathogène évolue dans le temps et dans l'espace, et iii) la description de la phylogéographie de ses populations permettra d'avoir une vue globale et de faire apparaitre des spécificités régionales et locales s'il y a lieu.

Avec un tel algorithme il sera possible de prédire l'endroit et probablement la vitesse avec laquelle la maladie de Lyme se propage, ce qui permettra *in fine* d'identifier les populations canadiennes à risque afin de mieux cibler les politiques de santé publique.

Par conséquent, notre étude est à considérer à différents niveaux du processus d'évaluation du risque dans un système de surveillance de la maladie de Lyme. Ces différents niveaux sont définis par les trois objectifs de cette étude présentés comme suit :

Objectifs

Pour ce travail, nous avons fixé 3 objectifs principaux :

1. Décrire les souches de *B. burgdorferi* présentes au Canada.
2. Investiguer les facteurs qui déterminent leur occurrence au Canada.
3. Mieux comprendre la dispersion de *B. burgdorferi* et établir un paysage génétique de la maladie de Lyme au Canada.

II. Synthèse des écrits

La synthèse de la littérature scientifique, décrivant les différentes composantes de la maladie de Lyme, permettra, en plus de fournir une synthèse des connaissances sur la BL, de comprendre les différents enjeux nationaux et internationaux de la santé publique dans lesquels s'inscrit ce projet de doctorat.

Le fil conducteur de l'interface homme-animal-écosystème est le rôle que jouent les changements environnementaux dans l'émergence et l'évolution des pathogènes en général et de *B. burgdorferi* en particulier. Dans une perspective globale, nous avons en premier lieu voulu situer le rôle des changements climatiques et environnementaux dans le paysage de la maladie de Lyme actuel et dans la dynamique de ce pathogène en Amérique du Nord. Nous allons ensuite présenter un aperçu de la diversité génétique de *B. burgdorferi* qui caractérise ce paysage. Pour comprendre l'importance de cette diversité sur le plan médical, nous exposons les différentes stratégies connues de ce pathogène pour infecter l'organisme. Enfin, nous décrivons le contexte épidémiologique actuel de la maladie de Lyme au Canada et en Amérique du Nord et les systèmes de surveillance mises en place pour prévenir cette maladie au Canada.

1. Phylodynamique de *B. burgdorferi*

Ce concept décrit les relations évolutives et les changements génétiques qui occurrent au sein d'une espèce et/ou dans une population en réponse à son environnement immédiat (Kurtenbach et al., 2006). Dans ce cadre, deux questions centrales se posent pour comprendre l'épidémiologie évolutive de ce pathogène: comment la diversité génétique de *B. burgdorferi* est générée et maintenue dans l'environnement ? Et comment la fluctuation des souches se produit entre différentes zones géographiques ?

1.1. La diversité génétique

À ce jour, seule l'espèce *B. burgdorferi* est connue responsable de la maladie de Lyme en Amérique du Nord (Ras et al., 1997; Seinost et al., 1999; Wang et al., 1999; Rudenko et al., 2013). Sa diversité génétique a été explorée dans plusieurs études (Wang et al., 1999; Bunikis et al., 2004; Anderson et Norris, 2006; Ogden et al., 2011, Margos et al., 2012; Hanincova et

al., 2013) par différentes approches moléculaires dont la précision de l'identification à l'échelle intra-spécifique est cruciale pour les études épidémiologiques, cliniques et évolutives (p.ex. typage avec les espaces ribosomaux *rrs-rrlA* et *rrfA-rrlB*, typage avec les séquences du gène codant pour la protéine *ospC*, typage avec MLST, typage avec le génome complet) (Wang et al., 2014). À titre d'exemple en Amérique du Nord, le génotypage avec des séquences de Multi-locus (MLST) a permis l'identification de 118 types de séquences (ST) aux profils alléliques différents (*pubmlst.org/bburgdorferi/*), contre 28 STs seulement connus en Eurasie (Jungnick et al., 2015). Des études récentes rapportent que même à l'échelle locale, *B. burgdorferi* présenterait une hétérogénéité génétique considérable (Rudenko et al., 2013). Au Canada, les connaissances des souches de *B. burgdorferi* sont limitées. Dans l'étude d'Ogden et al. (2011), les auteurs se sont intéressés à la diversité génétique de *B. burgdorferi* au Canada, mais les données de tiques provenant de la surveillance passive limite leur utilisation, car leurs origines géographiques sont incertaines (Ogden et al., 2006).

La dynamique de *B. burgdorferi* en Amérique du Nord est marquée par des périodes d'expansions et de contractions (Hoen et al., 2009), engendrant des phénomènes de spéciation tels que la radiation adaptative que connait actuellement ce pathogène dans cette partie du monde (Ogden et al., 2011; Margos et al., 2012). Ces phénomènes peuvent expliquer en partie l'importance de sa diversité génétique aux ÉU (Kurtenbach et al., 2002). En effet, durant des périodes d'expansion, les échanges génétiques (p.ex. recombinaison, mutations ponctuelles, transfert horizontal) sont fréquents entre les souches de *B. burgdorferi* (Mongodin et al., 2013).

1.2. Fréquence et changement de type de souches de *B. burgdorferi*

1.2.1. La sélection naturelle et le polymorphisme de niches multiples

En plus des phénomènes intrinsèques naturels et normaux (cités auparavant) qui contribuent à la variation génétique entre les souches et/ou entraînent l'évolution et l'émergence de nouvelles souches de *B. burgdorferi*, d'autres phénomènes extrinsèques exercent des pressions sur cette bactérie qui provoqueraient des fluctuations dans la distribution géographique de sa population (Ogden, 2015).

Plusieurs études, se basant sur des données de terrain, ont constaté que la grande majorité des allèles connus d'*ospC* sont géographiquement distincts (Ras et al., 1997; Wang et al., 1999; Qiu et al., 2002; Ogden et al., 2011; Rudenko et al., 2013). Aussi, des modèles de simulation, exploitant ce même gène à l'échelle du complexe bactérien de *B. burgdorferi* sl, ont permis de prédire une évolution adaptative rapide entre les génotypes à mesure que la taille de la population augmente (Mongodin et al., 2013). Par ailleurs, en suivant le processus de la sélection négative dépendamment de leurs fréquences, les génotypes rares peuvent être maintenus dans l'environnement, car le système immunitaire des hôtes aura tendance à exercer une grande pression sur les génotypes les plus fréquents (Brisson et Dykhuizen, 2004). En effet, le gène codant pour la protéine ospC est sous la pression de l'équilibrage de sélection (en anglais : balancing selection), notamment du système immunitaire des hôtes vertébrés (Rudenko et al., 2013), ce qui est probablement responsable du niveau élevé de polymorphisme de ce gène (Brisson et Dykhuizen, 2004). Dans ce contexte, un polymorphisme de niches multiples, notamment la disponibilité d'un grand éventail d'hôtes, aura un effet positif dans le maintien de la diversité de *B. burgdorferi* dans l'environnement (Radolf et Samuels, 2010). En Europe, plus de 240 espèces animales ont été signalées comme des hôtes pour les tiques et plusieurs parmi elles peuvent être des hôtes réservoirs potentiels de *B. burgdorferi* (Gern, 2008).

1.2.2. Changements environnementaux

a) Changements climatiques

Actuellement, il est établi que les changements globaux tels que les changements climatiques et les changements dans les habitats sont susceptibles d'avoir des impacts multiples et simultanés sur les écosystèmes (Ogden, 2015).

De nombreuses études ont souligné le rôle des changements climatiques dans l'apparition et la réapparition des maladies infectieuses notamment des maladies vectorielles. Ces bouleversements majeurs affectent la diversité du vecteur et des hôtes, leur abondance, leur dispersion, ainsi que le cycle de développement des agents pathogènes qu'ils hébergent, et la qualité de leurs habitats (Gubler et al., 2001). A titre d'exemple, les températures extrêmes augmenteraient le taux de mortalité des tiques (Ogden et al., 2013b), alors que le froid allongerait le cycle de développement des tiques en réduisant le nombre de larves qui

survivent pour devenir adultes (Ogden et al., 2014). Si le climat est trop froid, le cycle de vie sera trop long et les populations de tiques ne survivront pas (Lindsay et al., 1997 ; Ogden et al., 2005 ; Leighton et al., 2012). La saisonnalité peut aussi avoir un impact indirect sur le taux de transmission de *B. burgdorferi* en augmentant la probabilité de faire coïncider l'activité de recherche d'hôtes par les tiques avec l'activité des hôtes réservoirs (Ogden et al., 2005).

b) Changements des habitats

Le morcellement des milieux naturels et le changement de vocation des terres sont les principaux changements de l'habitat qui affectent indirectement et/ou directement le paysage de la maladie de Lyme en Amérique du Nord.

À titre d'exemple, la reforestation (c'est-à-dire la transformation des milieux agricoles en forêts) crée des habitats qui deviennent propices pour la tique et ses hôtes, particulièrement pour les cerfs (Wood et Lafferty, 2013). En effet, dans le Nord-est de l'Amérique du Nord, le reboisement dû à l'abandon des terres agricoles au cours du 20ème siècle aurait permis la recolonisation de ces milieux par le cerf et par conséquence l'expansion de la gamme des tiques. Ceci a engendré l'émergence de la maladie de Lyme vers la fin du 20^{ième} siècle dans cette partie du monde (Barbour et Fish, 1993). Par ailleurs, la fragmentation de l'habitat et la déforestation aux ÉU pourrait réduire la biodiversité des mammifères dans les forêts fragmentées, en favorisant les rongeurs réservoirs de *B. burgdorferi* (p.ex. souris à pattes-blanches) (Allan et al., 2003). Le fait que la biodiversité élevée pourrait agir comme un facteur protecteur, réduisant la transmission de *B. burgdorferi* grâce à l'effet de dilution (Ostfeld et Keesing, 2000), reste controversé, car la biodiversité peut causer l'effet inverse, à savoir l'amplification de la transmission de *B. burgdorferi* (Ogden et Tsao, 2009 ; Randolph et Dobson 2012).

Les changements environnementaux (habitat et climat) peuvent avoir des effets directs sur les souches de *B. burgdorferi*. En effet, en provoquant une expansion des populations des tiques et des hôtes, les souches de *B. burgdorferi* se disperseront au sein des milieux là où ces espèces existent déjà, mais également en dehors de ces milieux, si ces changements environnementaux engendrent une expansion plus large du vecteur et des hôtes. Ceci va provoquer une invasion de nouvelles zones géographiques par *B. burgdorferi* (Ogden et al., 20013a).

À court terme, la diversité de *B. burgdorferi* produite durant ce processus d'invasion aura un effet fondateur, et ensuite son adaptation aux populations d'hôtes naïfs provoquera une expansion rapide des souches de *B. burgdorferi* dans ces populations (Ogden et al., 2013a). À long-terme et si les changements environnementaux se stabilisent, *B. burgdorferi* peut évoluer en se spécialisant pour infecter des hôtes spécifiques. En effet, théoriquement la sélection naturelle favorise plutôt la spécialisation des parasites pour certains hôtes quand ces derniers présentent un développement stable, alors que les généralistes apparaissent quand l'évolution des hôtes est erratique, influencée par des changements majeurs (Combes 1997). Ceci permettra notamment le polymorphisme de niches multiples qui assureront la survie de *B. burgdorferi*. Ce processus a été déjà identifié en Europe où différentes espèces de *B. burgdorferi* sl survivent dans différentes niches écologiques. Ceci a un intérêt capital, car les manifestations cliniques montrées chez l'humain sont différentes pour les espèces adaptés pour différents hôtes (Kurtenbach et al., 2002; McCoy et al., 2003; Hanincova et al., 2006).

1.3. Phylogénie de *B. burgdorferi*

Pour des études évolutives intra-spécifiques, il est nécessaire d'avoir suffisamment de variation génétique entre les génotypes d'une même espèce. Cependant, le choix du ou des loci qui fournissent un bon niveau de polymorphisme est primordial pour capter un signal phylogénétique. À cet effet, plusieurs études ont ciblé différentes zones du génome de *B. burgdorferi* telles que les gènes codant des protéines de la surface externe (p.ex. ospC, ospA), l'espace inter-génique 16S-23S (*rrs-rrl*), et des gènes de ménages tels que recA et flagellin B (flaB) (Fukunaga et al., 1996 ; Valsangiacomo et al., 1997 ; Qiu et al., 2002 ; Bunikis et al., 2004; Hanincova et al., 2006).

Étant donné que le génome de *B. burgdorferi* est complexe (il renferme un long chromosome linéaire, des plasmides linéaires et d'autres circulaires), certaines études préfèrent combiner plusieurs loci (Postic et al., 2007; Margos et al., 2012; Hanincova et al., 2013), alors que d'autres utilisent un seul locus à la fois pour caractériser *B. burgdorferi* (Barbour et Garon, 1988; Liveris et al., 1995 ; Fukunaga et al., 1996 ; Lin et al., 2002). Si les loci proviennent de plusieurs plasmides, on aura tendance à favoriser le transfert horizontal de gènes pour expliquer la variation (Wang et al., 2000), alors que si les loci proviennent d'un même plasmide on risque de favoriser les évènements de recombinaison (Qiu et al., 2004).

L'étude de Margos et al. (2011) a souligné que dans le cas de *B. burgdorferi*, ces évènements sont probablement rares. Par conséquent, ces auteurs suggèrent plutôt l'utilisation du chromosome linéaire qui est plus adapté aux études sur les relations évolutives et démographiques, car les gènes de ménage sont plus conservés, plus stables et la variation génétique au niveau de ces gènes est proche du neutre.

Ainsi, l'utilisation des schémas multi-locus qui exploitent plusieurs gènes de ménages ont permis de distinguer clairement les populations Nord-Américaines et Européennes de *B. burgdorferi* (Margos et al., 2008). L'origine géographique de cette bactérie est sujette à la controverse, car l'étude de Ras et al. (1997) a prédit son origine en Amérique du Nord, alors que quelques années plus tard, Margos et al. (2008) ont prédit son origine en Europe. Cette différence est due aux types de marqueurs génétiques utilisés, car la première étude a exploité le gène codant pour ospC et la seconde étude a utilisé huit gènes de ménage. Récemment, en analysant en profondeur les variations au niveau de ces gènes de ménage, des études ont pu détecter au moins deux périodes importantes dans l'évolution de *B. burgdorferi* en Amérique du Nord. L'une s'est produite relativement récemment et l'autre il y a des milliers, voire des millions d'années (Hoen et al., 2009; Humphrey et al., 2010).

2. Épidémiologie contemporaine

2.1. Génotypage de *B. burgdorferi*

Les trois approches les plus utilisées pour typer *B. burgdorferi* sont: i) l'utilisation d'un seul gène chromosomique (p.ex. recA), ou ribosomal (p.ex. 16S), ou plasmidique (p.ex. ospC), ii) l'utilisation de plusieurs gènes ou multi-locus (p.ex. les gènes de MLST), et iii) l'utilisation du génome complet. Chaque méthode présente des avantages et des limites avec un gradient de perte d'information allant de l'utilisation du génome complet à un seul gène et un gradient inverse pour le coût économique (le coût de l'approche évolue avec la taille du génome ciblé).

2.1.1. Approche avec un seul locus

À titre d'exemple, nous présentons l'utilisation de l'espace inter-génique ou IGS 16S-23S (*rrs-rrlA*) et du gène codant pour ospC.

L'espace IGS couvre une région non-codante sur le chromosome. Il est fréquemment utilisé pour identifier des espèces de *B. burgdorferi* sl. Par contre, il présente une très faible

résolution pour capter une variation intra-spécifique (Margos et al., 2008). Son utilisation dans cette étude est motivée par le fait que les neuf types d'IGS connus pour *B. burgdorferi* se regroupent en trois groupes appelés RSTs (en anglais : « ribosomal sequence types» numérotés arbitrairement, RST1, RST2 et RST3). Chaque groupe est associé avec différents niveaux de pathogénicité (p.ex. RST1 est associé avec l'infection disséminée, alors que RST3 est associé avec l'infection localisée) (Wormser et al., 2005).

La protéine ospC est connue pour être impliquée dans le processus d'infection des hôtes par *B. burgdorferi*, puisque mécaniquement elle intervient directement lors du passage de la bactérie de la tique vers l'hôte (Tilly et al., 2006). Par conséquent, le gène du plasmide cp26 codant pour cette protéine de surface peut être l'un des déterminants génétiques nécessaires pour que l'infection se produise. Par ailleurs, son action reste limitée uniquement au stade précoce de l'infection, puisque après l'introduction de la bactérie dans l'organisme, cette protéine hautement antigénique disparait de la surface de la bactérie pour lui permettre d'échapper au système immunitaire (Skare et al., 2016). D'autres déterminants peuvent par la suite assurer la dissémination de *B. burgdorferi* à travers l'organisme, causant durant sa migration des inflammations et des réponses auto-immunes (Ogden et al., 2015a).

2.1.2. Approche avec plusieurs loci

Les méthodes à multi-locus sont considérées comme étant des techniques intermédiaires et elles restent actuellement préférables (Margos et al., 2008). En effet, l'approche MLST avec un schéma allélique de séquences de 8 gènes concaténées présente un pouvoir discriminatoire suffisant pour capter un signal phylogénétique, nécessaire pour investiguer une microévolution dans les souches de *B. burgdorferi* (Urwin et Maiden, 2003; Margos et al., 2008).

Des études génétiques ont permis de différencier des souches de *B. burgdorferi* qui varient par leur habilité à disséminer et à causer la maladie de Lyme (Schotthoefer & Frost, 2015). Récemment, Hanincova et al. (2013) ont mis en évidence chez l'humain le phénotype de pathogénicité différentielle que montrent des groupes de MLST-STs génétiquement proches nommés complexes clonaux (CCs). Par exemple, le CC37 est associé avec l'infection localisée, alors que le CC7 et le CC16 sont associés avec l'infection disséminée (Hanincova et al., 2013). Ceci suppose que la pathogénicité de *B. burgdorferi* est un trait complexe dont le

phénotype est régulé par l'interaction de nombreux déterminants génétiques qui demeurent méconnus (Ogden et al., 2015a). *B. burgdorferi* reste ainsi unique dans le monde bactérien, car jusqu'à présent on ne lui a pas attribué de facteurs de virulence et/ou des toxines susceptibles de causer la maladie comme le commun des bactéries (Hacker et Kaper, 2000).

2.1.3. Approche avec le génome complet

Le séquençage du génome complet et la détection des SNP (en anglais : single nucleotide polymorphism) ont permis des avancements considérables dans la compréhension des phénomènes génétiques qui surviennent lors des processus démographiques de beaucoup de pathogènes (Margos et al., 2008). Cependant, le coût élevé et les limites bio-informatiques rendent cette approche inaccessible pour des études de terrain et limitent son utilisation à de grandes échelles. Actuellement seul une dizaine de génomes complets de *B. burgdorferi* sont disponibles et rendus publiques (Schutzer et al., 2011).

2.2. Importance en santé publique de *B. burgdorferi*

2.2.1. Contexte phylogéographique actuel

Actuellement, *B. burgdorferi* est en pleine expansion aux États-Unis notamment dans les régions du Nord-est (NE) et du Midwest (MW). Au Canada, au début de la dernière décennie du siècle passé, seul le Long Point (Ontario) est connu pour abriter une population de tiques établie (Barker et al. 1988). Actuellement, *I. scapularis* est établie dans plusieurs régions du Sud : le Sud de Manitoba, le Sud d'Ontario, le Sud et le Sud-ouest de Québec et certaines régions des Maritimes (Ogden et al., 2008a), ce qui rend le risque de la maladie de Lyme accru dans ces régions (Ogden et al., 2013a).

L'image phylogéographique que montre *B. burgdorferi* avec de multiples expansions et contractions du Nord-est à l'ouest est différente de celle donnée par ces deux vecteurs : *I. scapularis* dans le Nord-est et le Midwest et le Sud des ÉU et *I. pacificus* dans l'ouest (Humphrey et al., 2010; Margos et al. 2012; Ginsberg et al., 2014). En effet, si la dispersion sur de longues distances du pathogène est attribuée aux oiseaux migrateurs transportant les tiques, la dispersion sur de courtes distances est surtout assurée principalement par des oiseaux sédentaires et des méso et micromammifères (Humphrey et al., 2010). Ceci suppose que l'expansion et la contraction des vecteurs ne sont pas totalement ou directement responsables

de la phylogéographie de *B. burgdorferi* (Margos et al. 2012). Les vecteurs anciens de *B. burgdorferi* pourraient être ainsi plutôt des spécialistes de rongeurs, et l'infection des rongeurs et des tiques nidicoles étaient probablement le principal moyen par lequel *B. burgdorferi* était dispersée. D'autant plus que la phylogéographie des principaux hôtes réservoirs (exemple : *Peromyscus*) en Amérique du Nord (Dragoo et al. 2006) semble pouvoir mieux expliquer la variation phylogéographique de l'agent pathogène (c'est-à-dire leurs territoires de distribution expliquent mieux ceux de *B. burgdorferi*). Ce patron a été clairement démontré en Europe (c'est-à-dire les structures de certaines espèces de *B. burgdorferi* suivent celles de leurs hôtes, p.ex. la structure de *B. afzelii* suit celle des petits mammifères, alors que celle de *B. garinii* et de *B. valaisiana* suivent celle des oiseaux) (Vollmer et al., 2011).

La présence de populations de *B. burgdorferi* de taille très réduite sur le territoire canadien (assez différentes des populations sources des ÉU) peut être un signal d'évènements fondateurs. Les données de surveillance passive au Canada montrent que globalement, ce sont les mêmes gammes de souches de *B. burgdorferi* qui arrivent sur le territoire canadien avec quelques exceptions de souches enregistrées au Canada, mais non signalées aux ÉU (Ogden et al., 2011). Elles sont transportées par les oiseaux migrateurs à partir des ÉU en suivant relativement le même schéma phylogéographique (c'est-à-dire ces souches sont très proches génétiquement et géographiquement) (Ogden et al., 2011). Par contre, des spécificités territoriales pouvaient apparaître, telles que la présence de populations de *B. burgdorferi* portant des allèles d'ospC différents (L et I) qui se produisent dans des régions différentes, à savoir le Sud-est et le Sud-ouest du Canada (Ogden et al., 2011). De même pour le vecteur *I. scapularis*, car l'analyse du gène Cox1 a révélé la présence de 3 haplotypes dans trois zones géographiques proches, le Sud et le Sud-ouest de Québec, et le Sud d'Ontario (Méchai et al., 2013).

2.2.2. Stratégies d'infection

Les multitudes stratégies de survie, d'infection des hôtes, d'invasion de l'organisme et de pathogénicité qu'utilise *B. burgdorferi* lui confèrent un statut unique parmi les spirochètes (Pulzova et Bhide, 2014). Elle est l'une des bactéries les plus étudiées, mais qui intrigue encore. Effectivement, une simple recherche avec *B. burgdorferi* comme mot clé dans PubMed, montre qu'il y a plus de 3500 publications scientifiques qui ont traité le sujet. Pour

comprendre le ‘’fitness’’ de cette bactérie, il est important de considérer son environnement immédiat, son vecteur et ses hôtes. En effet, *B. burgdorferi* est un parasite obligatoire mais extracellulaire dans l'hôte et le vecteur, adoptant différentes stratégies à différentes étapes de l'infection (Embers et al., 2004; Singh et al., 2004).

Le processus d'invasion de *B. burgdorferi* commence à partir des intestins de la tique où elle est attachée aux récepteurs TROSPA des cellules épithéliales grâce à ospA (Pal et al., 2004). Durant le repas sanguin, le pH et la température du milieu intestinal de la tique vont changer. Dès lors, *B. burgdorferi* entame sa migration vers les glandes salivaires à travers l'hémolymphe (Hoon-Hanks, 2012; Patton et al., 2012) en utilisant des interactions avec la lipoprotéine BBE31 et les récepteurs TRE31 (Zhang et al., 2011; Liu et Bonnet, 2014). Durant l'infection, *B. burgdorferi* surexprime la protéine ospC et utilise la protéine Salp15 des glandes salivaires pour faciliter sa transmission via la morsure de la tique vers l'épiderme de l'hôte (Wang et al., 2014). Après une période d'incubation durant laquelle *B. burgdorferi* se multiplie dans l'épiderme, elle utilise les enzymes protéolytiques de l'hôte (p.ex. plasminogène, MMP9) pour pénétrer et détruire la matrice extracellulaire et les jonctions intercellulaires (Gebbia et al., 2001; Floden et al., 2011). Une fois dans la circulation sanguine, simultanément *B. burgdorferi* augmente l'expression de la protéine de type facteur H (C-reactive protein binding proteins like factor H qui empêche les cellules C3 et C5 de s'attacher sur la surfaces de la bactérie) et réduit l'expression d'ospC sur ses surfaces pour échapper au système immunitaire (Xu et al., 2007; Gęca et al., 2016). Ainsi, la variation antigénique est l'une des stratégies qu'utilise *B. burgdorferi* durant l'invasion de l'hôte (Pulzova et Bhide, 2014).

La variation génétique des souches de *B. burgdorferi,* révélée par des techniques multi-locus (p.ex. MLSTs), semble également informer du degré de sévérité de la maladie de Lyme (Seinost et al., 1999; Dykhuizen et al., 2008; Hanincova et al., 2013). La diversité génétique de *B. burgdorferi* peut-être ainsi une clé pour comprendre son écologie (p.ex. associations hôte-pathogène) et ses phénotypes causant des maladies multi-systémiques (Qiu et Martin, 2014).

2.2.3. Pathogénicité différentielle

Celle-ci concerne les souches d'une espèce qui sont associées différemment avec la sévérité de la maladie. Pour *B. burgdorferi*, le mécanisme associé à la pathogénicité le plus connu est l'expression et la régulation de la protéine de surface C (ospC) (Schwan et al., 1995). Lors de la dissémination de la bactérie chez les mammifères sensibles à *B. burgdorferi*, cette protéine devient indispensable pour infecter l'hôte sur lequel la tique se nourrit (Stewart et al., 2006). En effet, la bactérie dans les intestins de la tique gorgée ou non favorise l'expression des protéines ospA et ospB. Lors de l'infection par contre, elle exprime plutôt la protéine ospC, qui sera dérégulée jusqu'à des niveaux indétectables une fois la bactérie dans l'organisme (Seemanapalli et al., 2010). Par ailleurs, plusieurs allèles ospC sont plus associés à la pathogénicité de *B. burgdorferi* (p.ex. I, A, H, K) que d'autres (p.ex. C, D, E, U) (Wormser et al., 2008). Ce phénotype de pathogénicité peut être favorisé par des associations de ce gène avec d'autres loci comme IGS (Bunikis et al., 2004; Grimm et al., 2004) ou aussi avec des complexes clonaux de STs, ce qui peut être important pour l'identification de certaines associations hôtes-pathogènes (c'est-à-dire la présence de déséquilibre de liaison peut être un signal d'associations préférentielles entre ces allèles dépendamment de l'espèce d'hôte) (Ogden et al., 2011; Margos et al., 2012).

2.2.4. Diagnostic différentiel

Dans le contexte d'émergence où la diversité génétique de *B. burgdorferi* est élevée, on peut s'attendre à ce que les tests sérologiques actuels ne soient pas suffisamment sensibles et spécifiques pour toutes les souches. En effet, les kits disponibles et largement utilisés sont calibrés pour une seule souche de *B. burgdorferi*, à savoir la B31 qui est proche de ST1 (Castillo-Ramírez et al., 2016). De plus, ces kits étant de classe II sur une échelle de 5 d'efficacité, l'interprétation des tests sérologiques peut être problématique (Johnson 2011). D'où la recommandation des agences de la santé publique du Canada et des ÉU d'utiliser deux tests complémentaires pour la sérologie de la maladie de Lyme. Le premier test est celui d'ELISA (Enzyme-Linked Immunosorbent Assay) ou IFA (immunofluorescence assay) qui est suivi d'un deuxième test de Western blot si le premier est positif (Wormser et al., 2006).

Le tableau clinique de la maladie de Lyme présente une centaine de symptômes dont certains sont spécifiques (p.ex. l'érythème migrant) et d'autres non-spécifiques (p.ex. fièvre,

maux de tête, fatigue). De plus, le problème des faux-positifs et des faux-négatifs est important du fait des faiblesses des tests disponibles (Brown et al., 1999 ; Ang et al., 2011), ce qui complique d'avantage l'établissement d'un diagnostic clair et rapide de la maladie de Lyme. À cet effet, l'ASPC soutient plutôt la complémentarité de l'approche de sérologie à deux tests et le diagnostic clinique qui tient compte des symptômes et de l'historique du patient (Agence de la Santé publique du Canada, 2012).

3. Contrôle et prévention de la maladie de Lyme

3.1. Système de surveillance de la maladie de Lyme au Canada

Brièvement, les politiques de surveillance de la maladie de Lyme actuelles au Canada se basent sur la déclaration des cas humains rapportés par les différents organismes de la santé (fédéraux et provinciaux) et sur la synthèse des données qui proviennent de deux types de surveillances du vecteur : la surveillance passive et la surveillance active.

Le système de surveillance des cas humains endémiques a pour but de confirmer l'expansion géographique du risque pour la maladie de Lyme au Canada, et de caractériser les changements dans le nombre de cas (Ogden et al., 2015).

La surveillance passive se base sur les tiques collectées et soumises par les praticiens de la santé (c'est-à-dire les tiques provenant des humains) ou les tiques collectées et soumises par les vétérinaires (c'est-à-dire les tiques provenant des animaux de compagnies). Ce système a produit des données intéressantes pour la santé publique (Leighton et al., 2012), mais il a des problèmes de spécificité géographique causée par les tiques dispersées par des hôtes non-résidents tels que les oiseaux migrateurs (Ogden et al., 2016). La surveillance active consiste à visiter systématiquement des sites et collecter des tiques dans l'environnement et sur des rongeurs (Bouchard et al., 2015). Cette méthode est effectivement la méthode étalon pour identifier l'occurrence des populations de tiques. Dans les deux cas, ces méthodes sont utilisées pour identifier le risque dans l'environnement et les populations humaines à risque. Ce qui permet ensuite de sensibiliser ces populations aux risques et les informer des méthodes de prévention. La surveillance du vecteur montre un risque de maladie de Lyme accru dans les régions du Sud Canadien limitrophes aux ÉU. L'établissement de la tique ne cesse de s'élargir,

favorisé notamment par les changements climatiques et environnementaux (Bouchard et al., 2015).

3.2. Évaluation du risque environnemental de la maladie de Lyme

L'évaluation du risque environnemental a pour but de mieux détecter et prédire les populations à risque et d'établir en conséquence des politiques de prévention. Elle consiste en général à identifier les types du risque, les zones d'occurrence du risque et les facteurs qui provoquent l'évolution spatio-temporelle du risque.

Les modèles d'évaluation du risque de la maladie de Lyme connus à date au Canada se basent sur des modèles de prédiction statistiques (Guerra et al. 2002; Brownstein et al. 2003; Leighton et al. 2012, Diuk-Wasser et al. 2012), des niches écologiques (Johnson et al. 2016) ou des simulations (Ogden et al. 2008). Le but de ces analyses du risque était la prédiction des régions actuellement à risque en utilisant des données de surveillance active, ou la prédiction des effets des changements climatiques plausibles sur l'évolution de la maladie de Lyme en Amérique du Nord dans les décennies à venir en utilisant des modèles mathématiques. Cependant, ces modèles ne tiennent pas compte des facteurs tels que les mouvements des hôtes réservoirs (Simon et al., 2014) et les relations hôtes-souches de *B. burgdorferi* qui peuvent être des éléments importants dans l'évaluation des changements de la répartition spatiale du risque chez l'humain.

En conclusion, grâce à ce projet, nous allons investiguer l'un des aspects les moins traiter par la littérature scientifique, à savoir la diversité et l'écologie évolutive de *B. burgdorferi* en Amérique du Nord pour tenter de dégager des hypothèses concrètes qui aideront à expliquer l'épidémiologie actuelle de la maladie de Lyme au Canada.

III. Chapitre I : Description des souches de *B. burgdorferi* au Canada

Complex Population Structure of *Borrelia burgdorferi* in Southeastern and South Central Canada as Revealed by Phylogeographic Analysis

S. Mechai,[a] G. Margos,[b,c,d] E. J. Feil,[e] L. R. Lindsay,[f] N. H. Ogden[a,g]

Groupe de Recherche en Épidémiologie des Zoonoses et Santé Publique, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, Canada[a]; Ludwig Maximilians Universität München, Department for Infectious Diseases and Zoonoses, Munich, Germany[b]; National Reference Centre for Borrelia, Oberschleissheim, Germany[c]; Bavarian Health and Food Safety Authority, Oberschleissheim, Germany[d]; Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom[e]; National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada[f]; Zoonoses Division, Public Health Agency of Canada, Saint-Hyacinthe, Québec, Canada[g]

Received 12 November 2014 Accepted December 2014
Accepted manuscript posted online 12 December 2014
Citation Mechai S, Margos G, Feil EJ, Lindsay LR, Ogden NH. 2015. Complex population structure of Borrelia burgdorferi in southeastern and south central Canada as revealed by phylogeographic analysis. Appl Environ Microbiol 81:1309–1318. doi:10.1128/AEM.03730-14.
Editor: C. R. Lovell
Address correspondence to N. H. Ogden, Nicholas.ogden@phac-aspc.gc.ca.
Supplemental material for this article may be found at http://dx.doi.org/10.1128 /AEM.03730-14.

Abstract

Lyme disease, caused by the bacterium *Borrelia burgdorferi* sensu stricto, is an emerging zoonotic disease in Canada and is vectored by the blacklegged tick, *Ixodes scapularis*. Here we used Bayesian analyses of sequence types (STs), determined by multi-locus sequence typing (MLST), to investigate the phylogeography of *B. burgdorferi* populations in southern Canada and the United States by analyzing MLST data from 564 *B. burgdorferi*-positive samples collected during surveillance. A total of 107 Canadian samples from field sites were characterized as part of this study, and these data were combined with existing MLST data for samples from the United States and Canada. Only 17% of STs were common between both

countries, while 49% occurred only in the United States, and 34% occurred only in Canada. However, STs in southeastern Ontario and southwestern Quebec were typically identical to those in the northeastern United States, suggesting a recent introduction into this region from the United States. In contrast, STs in other locations in Canada (the Maritimes; Long Point, Ontario; and southeastern Manitoba) were frequently unique to those locations but were putative descendants of STs previously found in the United States. The picture in Canada is consistent with relatively recent introductions from multiple refugial populations in the United States. These data thus point to a geographic pattern of populations of *B. burgdorferi* in North America that may be more complex than simply comprising northeastern, Midwestern, and Californian groups. We speculate that this reflects the complex ecology and spatial distribution of key reservoir hosts.

Introduction

Lyme disease, caused by the spirochete *Borrelia burgdorferi* sensu stricto (henceforth called *B. burgdorferi*), is continuing to emerge in the United States and is now emerging in southeastern and south central Canada due to the northward expansion of the range of the tick vector *Ixodes scapularis* (1). Recent studies have emphasized the scope of genetic diversity of *B. burgdorferi* (2,–4). For members of the *B. burgdorferi* sensu lato complex (to which *B. burgdorferi* belongs), diversity is likely to reflect a combination of historic patterns of geographic dispersion and associations or coevolution with different reservoir host species (4, 5). There are three recognized risk areas for Lyme disease in the United States: the Northeast (NE), the upper Midwest (MW), and the West (particularly California). Genetic differentiation of *B. burgdorferi* from these three regions has been noted and is likely due to geographic isolation by landscape features (4, 6–9). These populations have undergone recent expansions, but the available evidence also points to phylogenetically deeper and more complex patterns of expansion, contraction, and population mixing in the ancient past. Nevertheless, to date, studies have suggested an apparent spread from the Northeast through the Midwest to California (4, 6).

In order to recreate how *B. burgdorferi* populations in North America have expanded and contracted in the recent and deep past, it is necessary to first generate data concerning contemporary phylogeographic patterns. These data could then give insight into the ecological conditions underlying current population expansions, which in turn can help in the development of new management strategies. Given that past rapid climate changes are thought to have been key drivers of changes in population size and gene flow mediated by population movements (10) and that current rapid warming is thought to be driving range changes of *I. scapularis* and *B. burgdorferi* (11, 12), it is important to understand how current climate change and range spread may impact the diversity of *B. burgdorferi*.

Study of the diversity of *B. burgdorferi* in North America is of immediate diagnostic and clinical utility, with the recognition that the consequences of infection (mild self-limiting cutaneous or severe systemically disseminated infections) may differ with the infecting strain (13), and the strain may interact with patient genetic heterogeneity in determining clinical outcomes (14). Furthermore, strains vary in their capacity to elicit antibody responses in early infection that are detectable by current gold-standard serological tests (15). Recent studies of *B. burgdorferi* diversity by multi-locus sequence typing (MLST), which uses housekeeping genes with neutral variation, suggested that lineages determined from these housekeeping genes predicted pathogenicity better than outer surface proteins expressed at the point of infection (13). This raises the hypothesis that pathogenicity in humans is a *B. burgdorferi* phenotype with possible origins in adaptation to different host species and geographic locations and may therefore be predictable.

We have previously analyzed, by MLST, the diversity of *B. burgdorferi* detected in ticks, which were collected from humans and domesticated pets in a passive tick surveillance program in Canada, and compared this diversity with that found in the United States (4, 16). This analysis suggested that the general longitudinal pattern of differentiation of strains occurring in the northeastern versus the upper Midwestern United States is reflected in strains seen in Canada, with some possible skewing of diversity due to founder events as *B. burgdorferi* invades (16, 17). However, while many of the ticks collected in this study were likely from Canada-resident *I. scapularis* populations, some were likely not, having been dispersed from the United States by migratory birds (4, 16). To get a better picture of the strain structure in Canada-resident *B. burgdorferi* transmission cycles, we performed MLST

analysis of *B. burgdorferi* in tick samples collected during active field surveillance in locations where *B. burgdorferi* is known to be locally transmitted by self-sustaining, reproducing *I. scapularis* populations in Canada and compared these sequences to those previously obtained in the United States.

Materials and Methods

Samples used in the study

The 107 Canadian samples characterized by MLST as part of this study were *B. burgdorferi* positive by PCR (16) and were recovered from ticks obtained by drag sampling in areas in Canada where *B. burgdorferi* is endemic or by collection from wild rodent hosts. Areas of endemicity are defined as locations where *B. burgdorferi* is being transmitted among wild-animal reservoirs by reproducing populations of *I. scapularis* ticks (Table 1 and Fig. 1). The methodology of field sample collection was previously described (3). The samples were collected mostly contemporaneously (2006 for samples from the Maritime Provinces [MR] and Manitoba [MB], 2007 to 2010 for samples from Quebec, and 2010 for samples from eastern Ontario), but the oldest samples were those archived from collections conducted in 2001 at Long Point, Ontario (ONLP). DNA was extracted from ticks and screened for *B. burgdorferi* infection by PCR, as previously described (18). PCR-positive ticks were then used for MLST analysis, as previously described (19). Briefly, MLST was conducted by nested PCR for each of the eight housekeeping genes (clpA, clpX, nifS, pepX, pyrG, recG, rplB, and uvrA), using HotStarTaq (Qiagen, Germany) as previously described (3). PCR fragments were sequenced in the forward and reverse directions and manually compared by using DNASTAR (Lasergene). Sequences of new sequence types (STs) and new alleles were submitted to (and are available from) the mlst.net database (http://borrelia.mlst.net/). To reduce the likelihood of amplification of sequences from different strains coinfecting samples, we prescreened samples for mixed infections by amplifying the chromosomal rrs-rrlA (16S-23S) intergenic spacer (IGS) region, as previously described (3), and then sequencing the amplicons. Any rrs-rrlA amplicons that revealed ambiguous (i.e., two or more equally plausible bases at one position) bases or sequences upon examination of sequence traces were considered to suggest that the

samples contained possible mixed infections and were not subject to MLST analysis. The rrs-rrlA sequence was chosen for this purpose because it is one of the most variable sequences of *B. burgdorferi* used for strain analysis (20). Also, any samples that yielded any other ambiguous sequences of the housekeeping genes were not used for analysis. This led to the rejection of 60 samples (48 on the basis of IGS sequences and 12 on the basis of housekeeping gene sequences).

Table 1 : Field sites in Canada where ticks were collected from the environment or captured rodents at locations where tick populations have been established

Province	Location	No. of sites	No. of samples	Type of sample[a]
MB	Lake of the Woods, northwest shore	1	32	Questing ticks (19 AM, 13 AF)
ON	Long Point Provincial Park	1	7	Questing ticks (1 AM, 6 AF)
ON	Thousand Islands region	1	15	Questing ticks (4 AM, 4 AF, 7 N)
QC	Sites in Montérégie	11	35	16 questing ticks (4 AM, 11 AF, 1 N) 12 ticks from rodents (9 N, 3 L) 7 ticks from deer (1 AM, 6 AF)
NS	Lunenburg	1	22	15 questing ticks (10 AM, 3 AF, 2 N) 7 ticks from rodents (7 L)

aM, adult male; AF, adult female; N, nymph; L, larva.

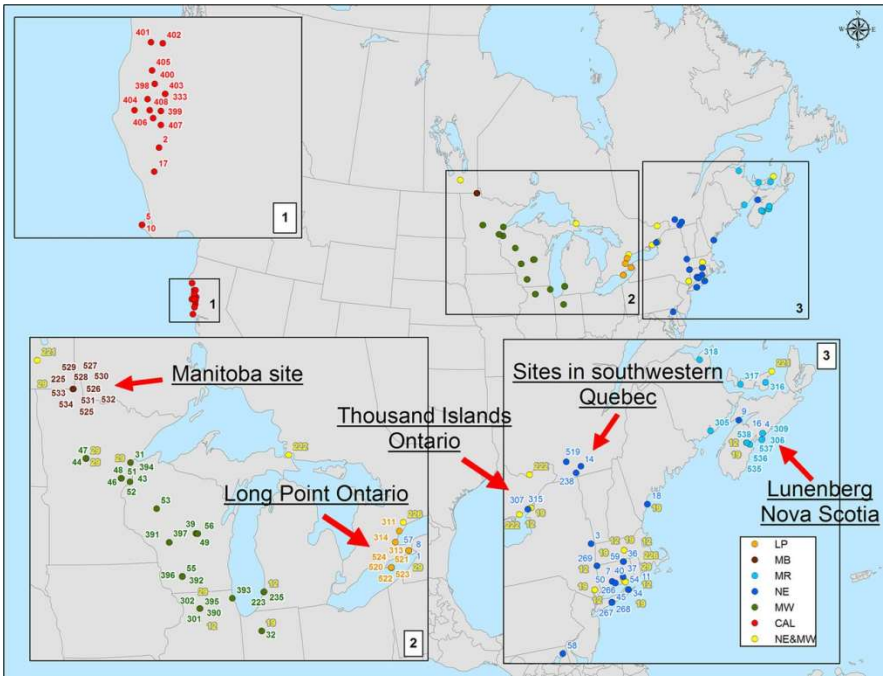


Figure 1 : Locations where *B. burgdorferi* samples used in this study were collected and the geographic distribution of different MLST STs of *B. burgdorferi* used in this study.
The colored points indicate locations where all samples analyzed in the study were collected, while red arrows indicate the locations of field sites where new samples in this study were obtained. The different-colored points correspond to STs found in different geographic regions. Colors: cyan, STs found only in the Maritimes (MR); orange, STs found only at Long Point, Ontario (ONLP); brown, STs found only in Manitoba (MB); blue, STs found across the Northeast (including Quebec, the Thousand Islands region of Ontario, the Maritimes, and the northeastern United States) (NE); green, STs found in the Midwestern United States (MW); yellow, STs found in northeastern and Midwestern locations of the United States and Canada (NE+MW); red, STs found only in California. (Maps were created using ArcGIS 10.1.)

For many analyses, an additional 4 samples collected during field surveillance in Quebec (3) and 20 samples collected during passive tick surveillance in Canada were also used because they carried novel STs that have to date been found only in Canada (see Table 1 in reference 4). The sequences of these samples were obtained from the mlst.net database.

For phylogenetic analyses, Canadian samples were combined with an additional 453 samples from the United States (all that were available at the time of analysis), the sequences of which were also obtained from the mlst.net database (http://borrelia.mlst.net/).

Nucleotide sequence analysis

New sequences obtained in Canada were compiled in Lasergene (DNAStar, Madison, WI,

USA), and these STs and their individual alleles were compared against existing STs and alleles in the mlst.net database by using the Single Locus Sequence Query and Allelic Profile Query functions. New alleles and STs were allocated identification numbers according to the protocol of the mlst.net database. The geographical distribution of STs was visualized using ArcGIS version 10.2 (ESRI). Population diversity (number of STs per sample) in each study site (in Canada) or region (in the United States) was calculated by dividing the number of individual STs at a site or region by the number of samples collected at that site/region. The diversity of STs in different sites and regions was identified in this process for comparison of populations and mapping.

Regarding the frequency of any new STs discovered in Canadian locations/regions, our null hypothesis was that the prevalence of newly discovered STs in the Canadian samples was not significantly different from the upper 95% confidence interval (CI) for the prevalence in the regions of the United States to the south that are the most likely source populations for *B. burgdorferi* carried northwards by migratory birds or other hosts (the upper Midwest for samples from Manitoba and Long Point and the Northeast for samples from eastern Ontario, Quebec, and the Maritimes). Therefore, the prevalence of any novel STs in Canadian locations was compared to the prevalence in the NE (0/363; 95% CI = 0 to 0.01) or the MW (0/62; 95% CI = 0 to 0.06) by Fisher's exact test.

Genetic diversity, phylogenetic relationships, and population structure

We analyzed the genetic diversity and population structure of *B. burgdorferi* strains from Canada and compared the results from Canadian samples with those from samples from the United States, using a number of analyses. Phylogenetic relationships were reconstructed by using MrBayes v3.2.1 software (21), in which Markov chain Monte Carlo (MCMC) samplings were run for 500,000 generations, with trees being sampled every 1,000th generation (22).

Pairwise F_{ST} values were calculated for the *B. burgdorferi* populations in different regions/locations by using the ARLEQUIN 3.1 program (23), with 100 permutations being run to assess the significance of the FST value. The level of significance was altered from a P value of <0.05 by Bonferroni correction to a P value of <0.001 to account for multiple pairwise comparisons.

Allelic profiles were analyzed by using eBURST (24) and global optimal eBURST

(goeBURST) (25). eBURST is based on a simple model of clonal expansion and divergence and provides a convenient method to establish relationships of descent for bacterial populations. goeBURST allows a global optimization procedure (instead of local optimization), an extended set of tiebreak rules, and improved graphical representation of clonal complexes, including double-locus variants (DLVs) and triple-locus variants (TLVs). Both algorithms are tailored for the use of MLST data and cluster STs as disjointed tree collections based on a set of hierarchical rules related to the number of single-locus variants (SLVs), DLVs (eBURST), and TLVs (goeBURST). The minimum number of identical loci for group definition was set to 5, and the minimum count of SLVs for subgroup definition was set to 0. The same samples were used in goeBURST to obtain a graphical display of clonal complexes. The Minimum Spanning Tree extension of PHYLOVIZ V1.0 (26) was used to visualize the possible evolutionary relationship between STs according to their allelic profiles in the goeBURST diagram. A bootstrap procedure implemented in eBURST gave statistical confidence to the assignment of clonal complex founders, which were inferred as the ST within a clonal complex that had the highest number of single-locus variants.

The population structure of the different STs identified across the United States and Canada was computed with Bayesian Analysis of Population Structure (BAPS) version 6.0 (27), using clustering with a linked locus module and codon model as recommended for MLST data. In this process, mixture analysis was performed with K values from 2 to 20, and optimal partitions were identified by the maximum log marginal likelihood value.

Investigation of the degree of recombination among B. burgdorferi strains and admixture among populations

The relative contribution of recombination (r) and mutation (m) to variation among the sequences was estimated by the r/m ratio calculated with ClonalFrame software v1.1 (28). The r/m value was obtained with 50,000 burn-in iterations, followed by 50,000 MCMC iterations and a thinning interval of 100 iterations before recording the parameter values for the posterior sample. The initial value for m was Watterson's theta value calculated for the sample by using DnaSP5 (29).

To estimate the contribution of admixture to genetic variation among BAPS groups, admixture analysis was conducted with BAPS 6.0 using the following parameters: a minimum population

size of 3, 100 iterations used to estimate the admixture coefficient for individuals, 200 reference individuals from each population, and 20 iterations used to estimate the admixture coefficient for reference individuals. Gene flow among the populations was plotted in BAPS 6.0, which uses a model-based representation of the molecular variability of populations and their affinities toward each other (30).

Results

Nucleotide sequence analysis

A total of 131 samples from Canada were used in the analysis, of which 111 were collected from field sites where *B. burgdorferi* is now endemic. Sequences of four of these samples collected in the field were available from a previous study (3). The 107 samples characterized by MLST as part of this study comprise 39 STs, 21 of which were novel (new STs were assigned the numbers 225 and 519 to 538) (see File S1 in the supplemental material). Two novel alleles were identified, both from ticks collected at Long Point, Ontario. These alleles corresponded to recG allele 167 for ST522 and clpA allele 182 for ST524. There were 20 samples from Canada collected during passive surveillance that had STs found only in Canada.

Genetic diversity and geographic distribution of STs

The 564 *B. burgdorferi* samples used for analyses in this study were divided into 111 STs. Of the STs already in the mlst.net database, 18 STs were unique to California, 27 STs were unique to the Midwest (6 of which have been found in "Midwestern" Canada [14]), 29 STs were unique to but widespread in the Northeast (including 14 found in eastern Canada, from eastern Ontario to the Maritimes [14]), and 6 STs occurred in both the Midwest and the Northeast (Table 2 and Fig. 1). In the samples characterized for the first time in this study, 39 STs were found. Twenty-one STs were unique to Canada, of which 4 STs were found only in the Maritimes (among 22 samples from Lunenburg, Nova Scotia); 5 STs were found only at Long Point, Ontario (from 7 samples); and 11 STs occurred only at the site in southeastern Manitoba (from 32 samples). In contrast, only one new ST was found in the samples from southern Quebec (from 35 samples), and no new STs were found in the samples from the

Ontario Thousand Islands site (from 15 samples). In total, including STs identified in previous studies, 54 STs are unique to the United States, 38 STs are unique to Canada, and only 19 STs are common to both countries. Thus, there were STs that occur across wide regions (those occurring in California and those occurring across the northeastern United States, the upper Midwestern United States, or both the northeastern and upper Midwestern United States and Canada) and STs that that to date have been found only in specific sites in Canada (those at Long Point, Ontario; Lunenburg, Nova Scotia; and the field site in Manitoba) (Fig. 1 [note that this map also shows Canada-specific STs from samples obtained during passive surveillance]).

Table 2 : New and total STs in samples collected from sites in Canada [a]

Site of sample collection	Total no. of samples	Total no. of sites	Total no. of STs [b]	No. of STs per sample	No. of STs unique to location	No. (proportion) of samples carrying unique STs	No. of unique STs per sample
Field sites in Canada							
MR	22	1	12	0.54	4	4 (0.19)†	0.18
QC	35	11	10	0.29	1	1 (0.03)	0.03
ONTH	15	1	7	0.47	0	0	0
QCTH	50	14	12	0.24	1	1 (0.02)	0.02
"NE" Canada	72	15	17 (8, 3, 0)	0.24	5	5 (0.07)	0.07
ONLP	7	1	7	1	5	5 (0.71)†	0.71
MB	32	1	19	0.59	11	13 (0.41)†	0.34
"MW" Canada	39	2	24 (0, 1, 6)	0.61	16	18 (0.46)	0.41
Sites in USA from mlst.net database							
NE	363	30	29	0.08			
MW	62	23	30	0.48			
California	28	23	18	0.64			

[a] Data on STs from the United States already in the mlst.net database are shown by geographic region for comparison. MR, Atlantic Maritime Provinces; QC, Quebec; ONTH, eastern Ontario's Thousand Islands; QCTH, Quebec and Thousand Islands combined; ONLP, Long Point, Ontario; MB, Manitoba; NE, northeastern United States; MW, Midwestern United States. "NE" Canada comprises data from the Atlantic Maritime Provinces, Quebec, and eastern Ontario's Thousand Islands combined, while "MW" Canada comprises data from Long Point, Ontario, and Manitoba combined. † indicates that the prevalence of samples carrying unique STs was significantly ($P < 0.001$) greater than the upper 95% confidence interval for their possible prevalence in source locations in the United States.

[b] Numbers in parentheses indicate the numbers of STs that were previously found in the northeastern, the Midwestern, and both the northeastern and Midwestern United States, respectively.

In the samples from the United States, the level of ST diversity (the number of STs per sample) was higher in the Californian samples (0.64; 18/28 samples) than in samples from the upper Midwest (0.48; 30/62) and lowest in samples from the Northeast (0.08; 29/363) (Table

2). Canadian samples from Manitoba and Long Point in Ontario, locations which correspond to the same longitude as the upper Midwestern United States, also showed a higher number of STs/sample (0.61; 24/39) than did samples from Ontario's Thousand Islands, Quebec, and Lunenburg, Nova Scotia (0.24; 17/72), which correspond to the same longitude as the northeastern United States (Table 2). STs from Lunenburg, Nova Scotia, were, however, more diverse than the combined samples from eastern Ontario and southern Quebec, with 0.54 and 0.24 STs per sample, respectively (Table 2). The prevalence of STs unique to the Manitoba site and to Long Point, Ontario (13/32 and 5/7 samples, respectively), was significantly higher than the upper 95% CI for their possible prevalence in the upper Midwestern United States ($P < 0.001$ for both). The prevalence of STs unique to Lunenburg, Nova Scotia (4/22 samples), was significantly higher than the upper 95% CI for their possible prevalence in the northeastern United States ($P < 0.001$). However, the prevalence of the ST unique to Quebec (1/35) was not significantly higher than the upper 95% CI for its possible prevalence in the northeastern United States ($P > 0.08$).

STs previously recorded in the northeastern United States that were also found in Canadian sites were found only in eastern Canada (from Thousand Islands, Ontario, eastwards). STs previously recorded in the upper Midwestern United States that were also found in Canada were found only in the more western Canadian sites (Long Point, Ontario, and the Manitoba site). On the basis of these observations of differences in STs among sites and regions, below, we keep the notation NE for STs found in the northeastern United States (although some STs that are found in the northeastern United States are also found in eastern parts of Canada) and MW for STs found in the upper Midwestern United States (although some STs that are found in the upper Midwestern United States are also found in western Ontario and Manitoba). For Canada, we consider the following sites/regions as comprising different ranges of STs: the site in Manitoba (MB), Long Point, Ontario (ONLP), and the Maritimes (MR). Given the similarity in STs and in their diversity, STs from the Thousand Islands region of Ontario and the sites in southern Quebec were considered one group, QCTH.

Phylogenetic relationships among STs

In the phylogenetic tree, each clade frequently comprised members of each broad geographic

region of North America (Fig. 2), suggesting ancient genetic signatures. For example, while the STs from California remained absent in all other regions, some of them (ST2 and ST403) are closely related genetically to ST1 from the northeastern United States and are in the same clade. New MB, ONLP, and MR STs were spread over different clades in the phylogeny. The new QCTH ST (ST519 from sites in Quebec) was most closely related to ST316 from, and unique to, MR (Fig. 2). Pairwise F_{ST} values (Table 3) supported moderate genetic differentiation and population structuring among the different geographic regions in the United States, and also, MB STs showed the highest value (F_{ST} = 0.19164) for genetic differentiation from the NE STs (Table 3).

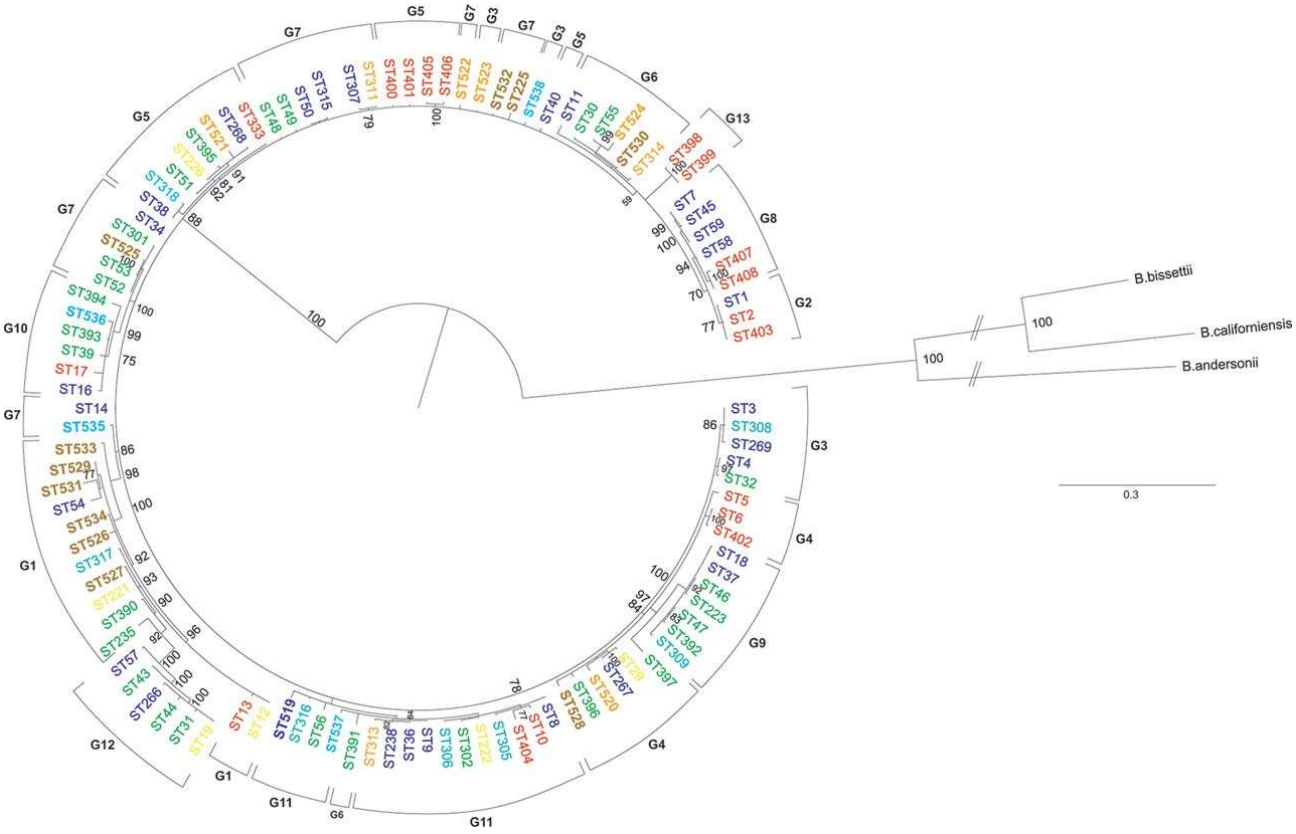


Figure 2 : Bayesian phylogenetic tree for the 111 STs of *B. burgdorferi*.
STs are color coded according to their geographic location. Cyan, Maritime Provinces; orange, Long Point, Ontario; brown, Manitoba; blue, northeastern United States, eastern Ontario, and southwestern Quebec; green, Midwestern United States; yellow, STs found in both the northeastern and Midwestern United States; red, California. Outgroups used to root the tree were *Borrelia californiensis*, *B. andersonii*, and *B. bissettii*. Posterior probabilities of >70% are shown beside the nodes. The bar corresponds to the number of substitutions per unit

branch length. Membership of STs in one of 13 elucidated BAPS groups is also indicated on the tree.

Table 3: Matrix of pairwise FST values of the STs in different geographic regions of North America [a]

	Pairwise F_{ST} value						
Region	**NE**	**MW**	**QCTH**	**MR**	**LP**	**MB**	**California**
NE	0						
MW	**0.10094**						
QCTH	0.08202	0.03131					
MR	0.05961	0.02552	0				
ONLP	0.07614	0.01173	0.00618	0.02804			
MB	**0.19164**	0.0652	0.10063	0.08337	0.07865		
California	**0.0758**	**0.06103**	0	0.02246	0.03457	**0.08927**	0

[a] Values in boldface type are significant at a threshold for significance of $\alpha = 0.0011$.

B. burgdorferi population structure

The goeBURST algorithm revealed that the *B. burgdorferi* STs are divided into 18 clonal complexes when only SLVs were included or 16 clonal complexes when DLVs were included as well as 45 singletons (Fig. 3). Except for ST403, ST2, and ST13, which formed a clonal complex with STs from outside California (ST1 and ST12), the remaining Californian STs formed clonal complexes with local STs, or they were singletons. While novel STs from MR were unique to that location, they were mostly SLVs, DLVs, or TLVs of STs found in the northeastern United States. Nearly all STs from field sites and passive surveillance in QCTH were the same as or closely related (SLV) to NE STs. The exception, ST519, was a DLV of an ST found to date only in the Maritimes (see the section above). STs from ONLP formed clonal complexes mostly with NE STs (Fig. 3) but appeared to have ancestry from both NE and MW STs, being SLVs of STs found in the northeastern United States and DLVs or TLVs of STs found in the Midwestern United States. STs from MB were divergent from MW STs but were most frequently SLVs, DLVs, or TLVs of STs found in the Midwestern United States. While the MB, MR, and ONLP STs diverged from U.S. STs and those from QCTH, they were most closely related to, and often had likely ancestors in, STs found immediately to the south: MB with the MW STs and MR with the NE STs. However, STs from ONLP, which lies on longitude 80°W, which nowadays separates NE and MW populations (14), had elements of relation and ancestry to both MW and NE STs. Inferred founders with >60% bootstrap support

were STs from the northeastern United States for two clonal complexes; other inferred founder (or possible founder) STs had <60% bootstrap support, and these also included the NE STs, the MW STs, and both the NE and MW STs (Fig. 3).

BAPS analysis best supported the existence of 13 subpopulations for which the log marginal likelihood value was highest. These subpopulations were not geographically structured, which is consistent with the lack of geographic structuring on the basis of phylogeny. This often suggests an ancient population structure and/or relatively high migration rates, although the latter is unlikely for *B. burgdorferi* in North America (4). BAPS groups 5 and 7 had STs from all seven geographic locations and contained 13 and 17 STs, respectively. BAPS groups 1, 4, and 11 contained STs from 6 geographic regions with 13, 9, and 15 STs, respectively. Groups 6 and 10 had STs from 4 geographic regions and contained 8 and 5 STs, respectively. Groups 2, 8, and 12 contained 3, 6, and 6 STs, respectively, which occurred in 2 to 3 geographic regions. BAPS group population 13 comprised only STs from California (ST398 and ST399) (Fig. 2 and 3). 3). There was general concordance between BAPS groups and clonal complexes revealed by goeBURST. However, two BAPS groups contained members from more than one clonal complex, and singletons were divided among BAPS groups (Fig. 3). Admixture analysis by BAPS revealed significant admixture ($P < 0.05$).

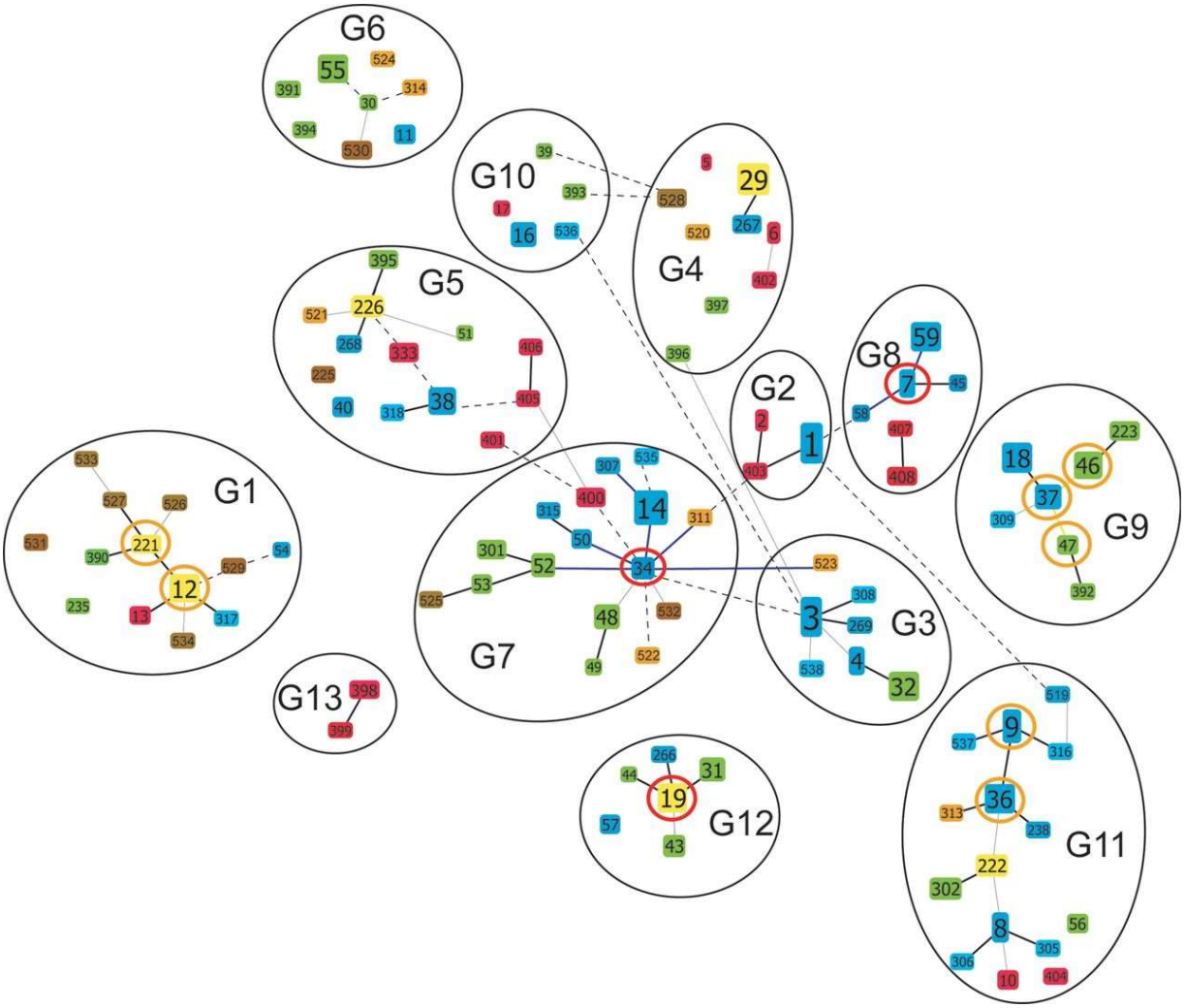


Figure 3: goeBURST network of the 111 STs of *B. burgdorferi* used or obtained in this study. STs are color coded according to their geographic location. Blue, northeastern United States, Quebec, the Thousand Islands region of Ontario, and the Maritimes; green, the Midwest; yellow, STs occurring in both the Northeast and the Midwest; red, California; cyan, STs occurring only in the Maritimes; orange, STs occurring only at Long Point, Ontario; brown, STs occurring only in Manitoba. Colored lines connecting STs in the network indicate the phylogenetic links between STs and the degree of support: black lines are inferred without tiebreak rules, blue lines are inferred by using tiebreak rule 1 (SLV), yellow lines are inferred by using tiebreaking at the frequency of the loci, and light gray lines are inferred by using tiebreak rule 2 (DLV). Inferred founder STs with >60% bootstrap support are circled in red, and potential founders with 35 to 60% support are circled in orange. TLVs are indicated by dashed lines. The population structure obtained by BAPS analysis is indicated as circles surrounding the clonal complexes and singleton STs.

Investigation of the degree of recombination and admixture analysis

The r/m value estimated for the sequences was 0.0162 (95% credibility interval, 0.0002 to

0.1497) when the initial m (i.e., Watterson's theta estimated in DnaSP5) was 28.64. This indicated a very low contribution of recombination to variation among sequences (31). Where recombination was found by admixture analysis, it occurred mostly (90% or more) within BAPS groups, as represented in the network by self-looping arrows (Fig. 4).

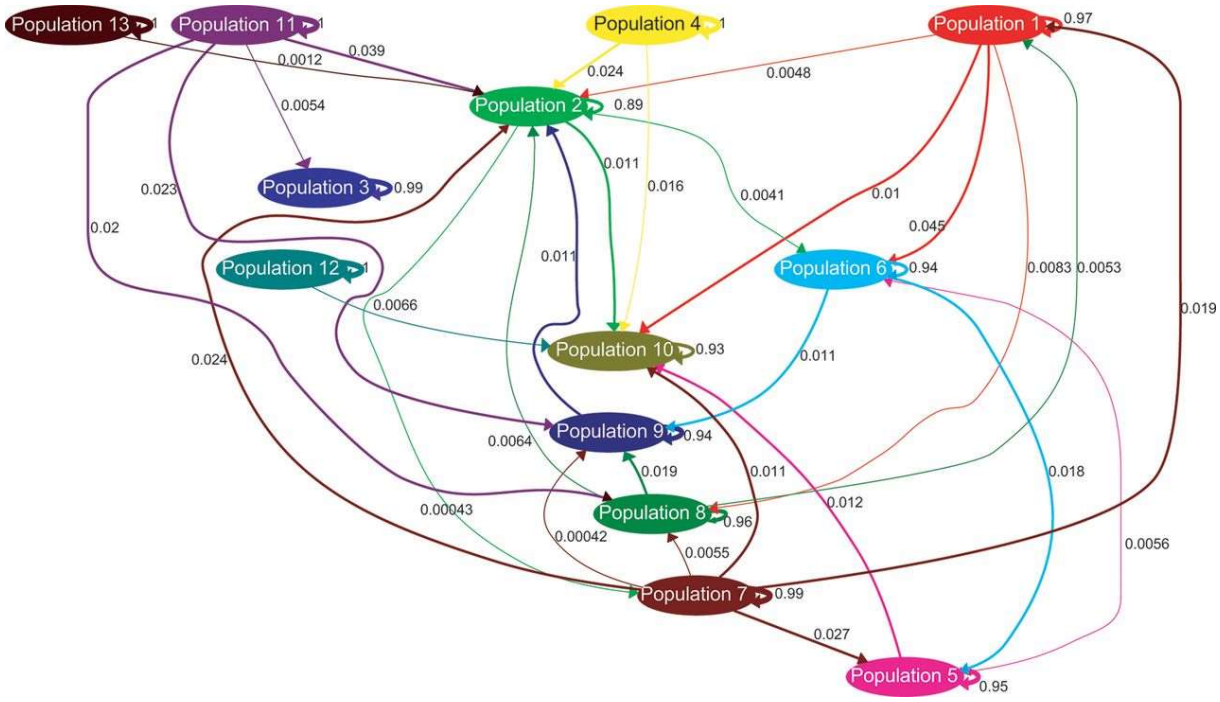


Figure 4: Gene flow occurring between different identified populations (BAPS groups), computed by using BAPS 6.0 software.
The arrows indicate the direction of gene flow, and the value accompanying each arrow represents the estimated average levels of DNA transition as relative gene flow weights among two or more populations. In this network, only significant admixture results ($P < 0.05$) are shown.

Discussion

Here we analyzed the STs of *B. burgdorferi* occurring in locations in Canada where *I. scapularis* tick populations are known to have become established (mostly in recent years), and this is the first cataloguing of STs from these locations. We had expected that we would find a range of STs in each site that originated from source populations in the United States directly to the south (having been carried in by northward-migrating birds or other hosts), with perhaps some skewing of the frequencies due to founder events (17). We did find STs that had been found previously in the United States in each site, and indeed, these STs have been found

mostly in locations directly to the south of the Canadian sites, with those in MB and ONLP having been found in the upper Midwestern United States and those in sites further east in Canada having been found in the northeastern United States. However, surprisingly, we found that in the whole North American data set, only approximately one-fifth of STs were common to both countries. One-half of the STs occurred only in the United States, and about one-third occurred only in Canada. Below, we discuss how our findings may improve our understanding of the phylogeography of *B. burgdorferi* in North America and, in the light of our analyses, raise hypotheses for the phylogeographic pattern observed in Canada.

Our study advances the understanding of the phylogeography of North American *B. burgdorferi* in general through (i) phylogenetic, goeBURST, and BAPS analyses; (ii) estimation of FST values among populations; (iii) reassessment of the origin of inferred clonal complex founders; and (iv) investigation of the small amount of variation among the housekeeping genes that was due to recombination. Since the description of *B. burgdorferi* in the northeastern United States in the 1980s (32), a clear pattern of geographic distribution of vector ticks, Lyme disease cases, and *B. burgdorferi* has been observed, with conspicuous gaps occurring between the Northeast and Midwest (in the region of Ohio) and California (33,–35). This pattern of apparent population structure has been attributed to geographic and other landscape barriers (4). Previous evaluations suggested that there have been multiple continent-wide population expansions (and presumably contractions) with local within-region population expansions in recent time (4, 6). This analysis confirms previous analyses showing that geographical patterns of ST occurrence in Canada and the United States are not represented in clades of the phylogenetic tree, the membership of clonal complexes by goeBURST analysis, and the membership of BAPS population groups: these groups contain STs from multiple geographic locations in most cases. Therefore, the underlying, ancestral genetic pattern is not geographically defined. If clades, clonal complexes, and BAPS groups of North American *B. burgdorferi* strains are not defined geographically (and assuming that mutations in different locations do not represent homoplasy), then perhaps they are defined ecologically. Previously, it was suggested that clonal complexes represent population expansions associated with introductions from Europe, with founder STs originating in the northeastern United States (6). However, in this study, potential founders were also STs that occurred in both the Northeast and Midwest, suggesting that the origin of the multiple

population expansions may be less clearly associated with introductions. An alternative hypothesis for the occurrence of clades and clonal complexes is that they represent broad associations with reservoir host species abundant at the time of past expansions and whose descendants persist today. There is recent evidence that some *B. burgdorferi* strains may be more efficiently transmitted by some host species (16, 36,–38), although there is no complete host specialization as there is for *B. burgdorferi* sensu lato species in Europe (39), and *B. burgdorferi* in North America remains a host generalist (40). Even so, FST values for comparisons between the northeastern, Midwestern, and Californian populations were lower than those among rodent-specialist *Borrelia afzelii* populations in Western Europe, which have a limited capacity for spatial mixing. However, these values were significantly greater than zero, which is not the case for the bird specialist *Borrelia garinii* in Europe, likely due to the considerable capacity for spatial mixing of bird-borne *B. garinii* populations (41). Thus, perhaps the North American populations show a pattern of spatial mixing that is at least partly driven by terrestrial host dispersions. Furthermore, we found that part of the genetic variation was associated with within-population horizontal gene transfer and recombination, as detected in previous studies (42). This has been observed to occur during infections of reservoir hosts (43). Almost all of the recombination occurred within BAPS groups, and this finding may support the idea that some degree of host association has driven the North American phylogeographic picture: both donor and recipient strains must have been capable of infecting the same individual host for a recombination event to occur. Further prospective studies to seek associations between host species and particular STs, clonal complexes, or clades are needed to support this hypothesis. Other hypotheses for the occurrence of the clades could, however, include population expansions and contractions associated with past climate changes (i.e., glacial and interglacial conditions [44]) and mass extinctions of vertebrate hosts (45).

Clustering analysis using goeBURST showed that STs unique to Canadian locations most likely had common ancestors with STs in U.S. regions immediately to the south of where they were found. A number of findings suggested that the *B. burgdorferi* populations in some of the Canadian locations (MB, ONLP, and MR) had characteristics that made them distinct from the known U.S. populations. First, the high proportion of unique STs in these Canadian locations, and the proportions of samples in each location that carried unique STs, precluded the idea that novel STs were found merely by chance due to extra field sampling effort increasing the

likelihood of finding additional rare STs. Second, F_{ST} values suggested moderate differentiation of the MB *B. burgdorferi* population from northeastern U.S. populations, and the FST value was higher than that for comparisons between U.S. populations for which barriers to gene flow have already been identified (4). Third, the ST diversity (as revealed by the number of STs/sample) in Canadian sites was greater than that in the corresponding regions (those to the south of them) in the United States. Together, these analyses suggest that it may not currently be possible to imply simple processes of invasion of strains occurring in MB, ONLP, and MR from the currently known *B. burgdorferi* populations in the northeastern and upper Midwestern regions of the United States. In contrast, however, STs from southern Quebec and eastern Ontario were almost all the same as those known to occur in the northeastern United States, which suggests that *B. burgdorferi* strains in these regions are direct invaders from the northeastern United States.

One hypothesis for the ST diversity observed in our study is that the MLST method used produces spurious STs by random or unpredictable amplification by PCR of different loci in samples carrying mixed-strain infections that were not detected by examination of traces. Although this cannot be ruled out in all cases, a prospective study showed that this was a very unlikely cause of the occurrence of new STs that were recombinations of previously identified alleles of MLST loci (see Files S2 and S3 in the supplemental material). A second hypothesis is that the *B. burgdorferi* populations in the Maritimes and western Ontario and Manitoba comprise refugial populations. From previous studies on refugial populations in other species, the expected observation would be that refugial populations comprise divergent strains that share one or a few ancestors (46). While there was some evidence of this pattern in the Manitoba site (BAPS group G1) (Fig. 3), it was clear that STs in each location (MB, ONLP, and MR) come from multiple populations and multiple parts of the phylogenetic tree and are derived from different ancestors. In fact, the diversity of strains in the Manitoba site was among the greatest of any location/region. Furthermore, while the *I. scapularis* population in Long Point could be refugial (47), local history suggests that ticks were only recently introduced into the sites in Manitoba and Nova Scotia (L. R. Lindsay, unpublished data). Therefore, the pattern of STs seen here would be more consistent with *B. burgdorferi* populations in these locations having been recently introduced from multiple populations in the United States. If so, then as the STs have not been discovered in the United States, these

STs may have originated in refugial source populations in the United States close to or bordering Canada where the ecology and diversity of *B. burgdorferi* have only recently begun to be explored (e.g., see references 48 and 49). This in turn suggests that the population structure of *B. burgdorferi* in North America is more complex than currently thought. We speculate that such a complex pattern may have arisen due to population expansions from postglacial refugia in northern regions of the United States that occurred with landscape change over the last century (50), combined with complex patterns of spatial mixing of B. burgdorferi strains. The equally complex phylogeographic patterns of key reservoir hosts such as *Peromyscus* species (51, 52) may have promoted differentiation in refugia that is now being amplified as changing ecological conditions increasingly support expansions of *B. burgdorferi* populations (53). The temporal aspects of sample collection, which spanned 2006 to 2010 for most sites but 2001 for one site, were not addressed in our study because of the assumption that such short time scales would not normally impact interpretations of sequences subject to the very low rates of mutation and recombination expected of housekeeping genes (19). However, it would be prudent in the future to investigate this assumption and explore rates of mutation and recombination in zones of emergence.

The main significance of our findings for Canada is that many STs at the western and eastern edges of the range of *I. scapularis* are different from those in the United States, while STs in Quebec and eastern Ontario are mostly the same as those already found in the northeastern United States. The ecological origins and consequences for pathogenicity of this pattern need to be further investigated. Our conclusion at present is that the apparent differentiation of populations of *B. burgdorferi* in Canada is most likely due to the importation of STs from refugia further south in the United States that have not been explored to date. However, further elucidation of the phylogeography of *B. burgdorferi* in North America and Canada, and of its ecological drivers, awaits more comprehensive and wider coverage of field sampling, culture, and isolation of strains; detailed analysis by whole-genome sequencing; and a better understanding of the mechanisms of *B. burgdorferi* dispersion.

Acknowledgments

This study was funded by the FQRNT and the Public Health Agency of Canada.

We gratefully acknowledge our colleagues Mike Drebot, Antonia Dibernardo, and Steve Tyler at the National Microbiology Laboratory, Public Health Agency of Canada, for sequencing of STs in this study. We also thank Yann Pelcat, Laboratory for Food-Borne Zoonoses, Public Health Agency of Canada, for creating Fig. 1.

References

1. Ogden NH, Lindsay LR, Sockett PN, Morshed M, Artsob H. 2009. The emergence of Lyme disease in Canada. CMAJ 180:1221–1224. http://dx .doi.org/10.1503/cmaj.080148.
2. Humphrey PT, Caporale DA, Brisson D. 2010. Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. Evolution 64:2653–2663. http://dx.doi.org/10.1111/j.1558-5646.2010 .01001.x.
3. Ogden NH, Bouchard C, Kurtenbach K, Margos G, Lindsay LR, Trudel L, Nguon S, Milord F. 2010. Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of Lyme disease risk emergence in Canada. Environ Health Perspect 118:909–914. http://dx.doi.org/10.1289/ehp.0901766.
4. Margos G, Tsao JI, Castillo-Ramírez S, Girard YA, Hamer SA, Hoen AG, Lane RS, Raper SL, Ogden NH. 2012. Two boundaries separate *Borrelia burgdorferi* populations in North America. Appl Environ Microbiol 78:6059 – 6067. http://dx.doi.org/10.1128/AEM.00231-12.
5. Vollmer SA, Feil EJ, Chu CY, Raper SL, Cao WC, Kurtenbach K, Margos G. 2013. Spatial spread and demographic expansion of Lyme borreliosis spirochaetes in Eurasia. Infect Genet Evol 14:147–155. http: //dx.doi.org/10.1016/j.meegid.2012.11.014.
6. Hoen AG, Margos G, Bent SJ, Diuk-Wasser MA, Barbour A, Kurtenbach K, Fish D. 2009. Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. Proc Natl Acad SciUSA 106:15013–15038. http://dx.doi .org/10.1073/pnas.0903810106.
7. Brinkerhoff RJ, Bent SJ, Folsom-O'Keefe CM, Tsao K, Hoen AG, Barbour AG, Diuk-Wasser MA. 2010. Genotypic diversity of *Borrelia burgdorferi* strains detected in *Ixodes scapularis* larvae collected from North American songbirds. Appl Environ Microbiol 76:8265– 8268. http: //dx.doi.org/10.1128/AEM.01585-10.
8. Barbour AG, Travinsky B. 2010. Evolution and distribution of the *ospC* gene, a transferable serotype determinant of *Borrelia burgdorferi*. mBio 1(4):e00153-10. http://dx.doi.org/10.1128/mBio.00153-10.
9. Brisson D, Vandermause MF, Meece JK, Reed KD, Dykhuizen DE. 2010. Evolution of northeastern and midwestern *Borrelia burgdorferi*, United States. Emerg Infect Dis 16:911–917. http://dx.doi.org/10.3201 /eid1606.090329.
10. Hofreiter M, Stewart J. 2009. Ecological change, range fluctuations and population dynamics during the Pleistocene. Curr Biol 19:R584–R594. http://dx.doi.org/10.1016/j.cub.2009.06.030.

11. Leighton P, Koffi J, Pelcat Y, Lindsay LR, Ogden NH. 2012. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. J Appl Ecol 49:457–464. http://dx.doi.org/10.1111/j.1365-2664.2012.02112.x.
12. Ogden NH, Lindsay LR, Leighton PA. 2013. Predicting the rate of invasion of the agent of Lyme disease *Borreliaburgdorferi*. J Appl Ecol 50:510– 518. http://dx.doi.org/10.1111/1365-2664.12050.
13. Hanincova K, Mukherjee P, Ogden NH, Margos G, Wormser GP, Reed KD, Meece JK, Vandermause MF, Schwartz I. 2013. Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. PLoS One 8:e73066. http://dx.doi .org/10.1371/journal.pone.0073066.
14. Iliopoulou BP, Guerau-de-Arellano M, Huber BT. 2009. HLA-DR alleles determine responsiveness to *Borrelia burgdorferi* antigens in a mouse model of self-perpetuating arthritis. Arthritis Rheum 60:3831–3840. http: //dx.doi.org/10.1002/art.25005.
15. Wormser GP, Liveris D, Hanincová K, Brisson D, Ludin S, Stracuzzi VJ, Embers ME, Philipp MT, Levin A, Aguero-Rosenfeld M, Schwartz I. 2008. Effect of *Borreliaburgdorferi* genotype on the sensitivity of C6 and 2-tier testing in North American patients with culture-confirmed Lyme disease. Clin Infect Dis 47:910 –914. http://dx.doi.org/10.1086/591529.
16. Ogden NH, Margos G, Aanensen DM, Drebot MA, Feil EJ, Hanincová K, Schwartz I, Tyler S, Lindsay LR. 2011. Investigation of genotypes of *Borrelia burgdorferi* in *Ixodes scapularis* ticks collected during surveillance in Canada. Appl Environ Microbiol 77:3244 –3254. http://dx.doi.org/10 .1128/AEM.02636-10.
17. Ogden NH, Mechai S, Margos G. 2013. Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. Front Cell Infect Microbiol 3:46. http://dx.doi.org /10.3389/fcimb.2013.00046.
18. Ogden NH, Lindsay RL, Hanincová 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 and of *Borrelia burgdorferi* and *Ana-plasma phagocytophilum* in Canada. Appl Environ Microbiol 74:1780– 1790. http://dx.doi.org/10.1128/AEM.01982-07.
19. Margos G, Gatewood AG, Aanensen DM, Hanincová K, Terekhova D, Vollmer SA, Cornet M, Piesman J, Donaghy M, Bormane A, Hurn MA, Feil EJ, Fish D, Casjens S, Wormser GP, Schwartz I, Kurtenbach K. 2008. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. Proc Nat lAcad Sci USA 105:8730 – 8735. http://dx.doi.org/10.1073/pnas.0800323105.
20. Bunikis J, Garpmo U, Tsao J, Berglund J, Fish D, Barbour AG. 2004. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. Microbiology 150:1741–1755. http://dx.doi.org/10.1099/mic.0.26944-0.
21. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, Larget B, Liu L, Suchard MA, Huelsenbeck JP. 2012. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol 61:539 –542. http://dx.doi .org/10.1093/sysbio/sys029.
22. Hall BG. 2011. Phylogenetic trees made easy: a how-to manual, 4th ed. Sinauer Associates, Sunderland, MA. Excoffier L, Laval G, Schneider S. 2005. Arlequin version

3.0: an integrated software package for population genetics data analysis. Evol Bioinform Online 1:47–50.
23. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186: 1518 –1530. http://dx.doi.org/10.1128/JB.186.5.1518-1530.2004.
24. Francisco AP, Bugalho M, Ramirez M, Carrico JA. 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics 10:152. http://dx.doi.org/10 .1186/1471-2105-10-152.
25. Francisco AP, Vaz C, Monteiro PT, Melo-Cristino J, Ramirez M, Carriço JA. 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13:87. http: //dx.doi.org/10.1186/1471-2105-13-87.
26. Corander J, Tang J. 2007. Bayesian analysis of population structure based on linked molecular information. Math Biosci 205: 19–31. http://dx.doi .org/10.1016/j.mbs.2006. 09.015.
27. Didelot X, Falush D. 2007. Inference of bacterial microevolution using multilocus sequence data. Genetics 175:1251–1266. http://dx.doi.org/10 .1534/genetics.106.063305.
28. Librado P, Rozas J. 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25:1451–1452. http://dx .doi.org/10.1093/bioinformatics/ btp187.
29. Tang F, Barbacioru C, Wang Y, Nordman E, Lee C, Xu N, Wang X, Bodeau J, Tuch BB, Siddiqui A, Lao K, Surani MA. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. Nat Methods 6:377–382. http://dx.doi.org/10.1038/nmeth.1315.
30. Vos M, Didelot X. 2009. A comparison of homologous recombination rates in bacteria and archaea. ISMEJ 3:199–208. http://dx.doi.org/10.1038 /ismej.2008.93.
31. Burgdorfer W, Barbour AG, Hayes SF, Benach JL, Grunwaldt E, Davis JP. 1982. Lyme disease: a tick-borne spirochetosis? Science 216:1317– 1319. http://dx.doi.org/10.1126/ science.7043737.
32. Diuk-Wasser MA, Gatewood AG, Cortinas MR, Yaremych-Hamer S, Tsao J, Kitron U, Hickling G, Brownstein JS, Walker E, Piesman J, Fish D. 2006. Spatiotemporal patterns of host-seeking *Ixodes scapularis* nymphs (Acari: Ixodidae) in the United States. J Med Entomol 43:166–176. http://dx.doi .org/10.1603/0022-2585(2006)043[0166:SPOHIS]2.0. CO;2.
33. Bacon RM, Kugeler KJ, Mead PS. 2008. Surveillance for Lyme disease. United States, 1992–2006. MMWR Surveill Summ 57:1–9.
34. Diuk-Wasser MA, Hoen AG, Cislo P, Brinkerhoff R, Hamer SA, Rowland M, Cortinas R, Vourc'h G, Melton F, Hickling GJ, Tsao JI, Bunikis J, Barbour AG, Kitron U, Piesman J, Fish D. 2012. Human risk of infection with *Borrelia burgdorferi*, the Lyme disease agent, in eastern UnitedStates.AmJTropMedHyg 86:320–327. http://dx.doi.org/10.4269 /ajtmh.2012. 11 -0395.
35. Brisson D, Dykhuizen DE. 2004. *ospC* diversity in *Borrelia burgdorferi*: different hosts are different niches. Genetics 168:713–722. http://dx.doi .org/10.1534/genetics.104.028738.
36. Derdáková M, Dudiòák V, Brei B, Brownstein JS, Schwartz I, Fish D. 2004. Interaction and transmission of two *Borrelia burgdorferi* sensu stricto strains in a tick-rodent

maintenance system. Appl Environ Microbiol 70:6783– 6788. http://dx.doi.org/10.1128/AEM.70.11.6783-6788 .2004.
37. Hanincová K, Ogden NH, Diuk-Wasser M, Pappas CJ, Lyer R, Fish D, Schwartz I, Kurtenbach K. 2008. Fitness variation of *Borrelia burgdorferi* sensu stricto strains in mice. Appl Environ Microbiol 74:153–157. http: //dx.doi.org/10.1128/AEM.01567-07.
38. Piesman J, Gern L. 2004. Lyme borreliosis in Europe and North America. Parasitology 129:S191–S220. http://dx.doi.org/10.1017/S0031182003004694.
39. Hanincová K, Kurtenbach K, Diuk-Wasser M, Brei B, Fish D. 2006. Epidemic spread of Lyme borreliosis, northeastern United States. Emerg Infect Dis 12:604 – 611. http:// dx.doi.org/ 10.3201/eid1204.051016.
40. Vollmer SA, Bormane A, Dinnis RE, Seelig F, Dobson AD, Aanensen DM, James MC, Donaghy M, Randolph SE, Feil EJ, Kurtenbach K, Margos G. 2011. Host migration impacts on the phylogeography of Lyme borreliosis spirochaete species in Europe. Environ Microbiol 13:184–192. http://dx.doi.org/10.1111/j.1462-2920.2010.02319.x.
41. Mongodin EF, Casjens SR, Bruno JF, Xu Y, Drabek EF, Riley DR, Cantarel BL, Pagan PE, Hernandez YA, Vargas LC, Dunn JJ, Schutzer SE, Fraser CM, Qiu WG, Luft BJ. 2013. Inter-and intra-specific pan-genomes of *Borrelia burgdorferi* sensu lato: genome stability and adaptive radiation. BMC Genomics 14:693. http://dx.doi.org/10.1186/1471-2164-14-693.
42. Rudenko N, Golovchenko M, Belfiore NM, Grubhoffer L, Oliver JH, Jr. 2014. Divergence of *Borrelia burgdorferi* sensu lato spirochetes could be driven by the host: diversity of *Borrelia* strains isolated from ticks feeding on a single bird. Parasit Vectors 7:4. http:// dx.doi.org/10.1186/1756-3305 -7-4.
43. Margos G, Vollmer SA, Ogden NH, Fish D. 2011. Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato. Infect Genet Evol 11:1545–1563. http://dx.doi.org/10.1016/j.meegid.2011.07.022.
44. Barnosky AD, Koch PL, Feranec RS, Wing SL, Shabel AB. 2004. Assessing the causes of late Pleistocene extinctions on the continents. Science 306:70–75. http://dx.doi.org/ 10.1126/ science.1101476.
45. Provan J, Bennett KD. 2008. Phylogeographic insights into cryptic glacial refugia. Trends Ecol Evol 23:564 –571. http://dx.doi.org/10.1016/j.tree .2008.06.010.
46. Barker IK, Surgeoner GA, McEwen SA, Artsob H. 1989. Distribution of the Lyme disease agent in wildlife reservoirs in Ontario. Can Dis Wkly Rep 15:141–142.
47. Hamer SA, Hickling GJ, Sidge JL, Rosen ME, Walker ED, Tsao JI. 2011. Diverse *Borrelia burgdorferi* strains in a bird-tick cryptic cycle. Appl Environ Microbiol 77:1999 – 2007. http://dx.doi.org/10.1128/AEM.02479-10.
48. Rydzewski J, Mateus-Pinilla N, Warner RE, Hamer S, Weng HY. 2011. *Ixodes scapularis* and *Borrelia burgdorferi* among diverse habitats within a natural area in east-central Illinois. Vector Borne Zoonotic Dis 11:1351– 1358. http://dx.doi.org/10.1089/vbz.2010.0160.
49. Wood CL, Lafferty KD. 2013. Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. Trends Ecol Evol 28:239 –247. http://dx.doi.org/ 10.1016/j.tree.2012.10.011.
50. WatersJH. 1963. Biochemical relationships of the mouse *Peromyscus* in New England. Syst Zool 12:122–133. http://dx.doi.org/10.2307/2411532.

51. Dragoo JW, Lackey JA, Moore KE, Lessa EP, Cook JA, Yates TL. 2006. Phylogeography of the deer mouse (*Peromyscus maniculatus*) provides a predictive framework for research on Hantaviruses. J Gen Virol 87:1997– 2003. http://dx.doi.org/10.1099/vir.0.81576-0.
52. Kurtenbach K, Hanincová K, Tsao J, Margos G, Fish D, Ogden NH. 2006. Key processes in the evolutionary ecology of Lyme borreliosis. Nat Rev Microbiol 4:660–669. http://dx.doi.org/10.1038/nrmicro1475.

Supplementary Material

Supplemental material for this article may be found at http://dx.doi.org/10.1128/AEM.03730-14.

Details of the sequence types (ST), individual loci, and source of samples collected in field surveillance in Canada (Table S1); supplemental text: assessment of the possibility that novel MLST STs are due to PCR anomalies of mixed-strain infections of field-collected samples (File S2); sequence traces of loci amplified from the different alleles of unmixed and mixed samples containing novel STs (Fig. S1).

Supplementary File 1

Supplementary Table 1. Details of the Sequence Types (ST), individual loci and source of samples collected in field surveillance in Canada

sample ID	ST	clpA	clpX	nifS	pepX	pyrG	recG	rplB	uvrA	region	Lat	Long	Source	Isolation Date	Sites
BH.2.55	1	1	1	1	1	1	1	1	1	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
L.30	1	1	1	1	1	1	1	1	1	MR	44.533	-64.499	Deer Mouse	7/16/2006	Lunenburg
S.185	1	1	1	1	1	1	1	1	1	ONth	44.351	-75.990	Environment	5/1/2010	Thousand Islands
S.423	1	1	1	1	1	1	1	1	1	ONth	44.351	-75.990	Environment	8/1/2010	Thousand Islands
S.426	1	1	1	1	1	1	1	1	1	ONth	44.351	-75.990	Environment	8/1/2010	Thousand Islands
S.506	1	1	1	1	1	1	1	1	1	ONth	44.351	-75.990	Environment	8/1/2010	Thousand Islands
19.1	1	1	1	1	1	1	1	1	1	QC	45.241	-72.968	White-tailed deer	6/1/2008	Farnham
28.4	1	1	1	1	1	1	1	1	1	QC	45.047	-73.239	White-tailed deer	11/1/2008	Saint Georges de Clarenceville
60.2	1	1	1	1	1	1	1	1	1	QC	45.274	-73.015	White-tailed deer	11/1/2008	Farnham
61.2	1	1	1	1	1	1	1	1	1	QC	45.167	-73.039	White-tailed deer	11/1/2008	Notre Damede Stanbridge
620	1	1	1	1	1	1	1	1	1	QC	45.308	-73.011	White-footed Mouse	9/1/2008	Farnham militaire
816	1	1	1	1	1	1	1	1	1	QC	45.308	-64.499	Environment	11/1/2008	Farnham militaire
F09.43	1	1	1	1	1	1	1	1	1	QC	45.308	-73.011	Environment	6/2/2009	Farnham militaire
BH.2	3	4	1	1	1	1	6	1	7	MR	44.533	-64.499	Environment	7/15/2006	Lunenburg
BH.2.27	3	4	1	1	1	1	6	1	7	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
BH.2.35	3	4	1	1	1	1	6	1	7	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
BH.2.5	3	4	1	1	1	1	6	1	7	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg

9															
CEM.1	3	4	1	1	1	1	6	1	7	MR	44.533	-64.499	Environment	7/15/2006	Lunenburg
S.340	3	4	1	1	1	1	6	1	7	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
S.433	3	4	1	1	1	1	6	1	7	ONth	44.351	-75.990	Environment	8/1/2010	Thousand Islands
S.487	3	4	1	1	1	1	6	1	7	ONth	44.351	-75.990	Environment	8/1/2010	Thousand Islands
33	3	4	1	1	1	1	6	1	7	QC	45.178	-73.346	White-footed Mouse	7/1/2008	Saint Valentin
35	3	4	1	1	1	1	6	1	7	QC	45.178	-73.346	White-footed Mouse	7/1/2008	Saint Valentin
154	3	4	1	1	1	1	6	1	7	QC	45.178	-73.346	Red squirrel	6/1/2008	Saint Valentin
225	3	4	1	1	1	1	6	1	7	QC	45.178	-73.346	Environment	5/1/2008	Saint Valentin
802	3	4	1	1	1	1	6	1	7	QC	45.226	-73.336	Environment	10/1/2008	Saint Blaise Surrichelieu
L.41	4	8	1	1	1	4	6	1	7	MR	44.533	-64.499	Deer Mouse	7/17/2006	Lunenburg
25.4	4	8	1	1	1	4	6	1	7	QC	45.058	-73.225	White-tailed deer	11/1/2008	Noyan
qc10.d.467	4	8	1	1	1	4	6	1	7	QC	45.783	-72.080	Environment	7/1/2010	Danville
S.243	8	5	5	4	5	5	5	1	6	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
85	8	5	5	4	5	5	5	1	6	QC	45.117	-73.212	Eastern chipmunk	6/1/2008	Henryville
F09.68	8	5	5	4	5	5	5	1	6	QC	45.308	-73.011	Environment	6/2/2009	Farnham militaire
BH.2.66	9	10	5	4	6	1	6	1	6	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
BH.2.7	9	10	5	4	6	1	6	1	6	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
L.50	9	10	5	4	6	1	6	1	6	MR	44.533	-64.499	Deer Mouse	7/18/2006	Lunenburg
61	9	10	5	4	6	1	6	1	6	QC	45.096	-72.968	White-footed Mouse	6/1/2008	Bedford
812	9	10	5	4	6	1	6	1	6	QC	45.308	-73.011	Environment	11/1/2008	Farnham militaire
BH.2.29	12	3	3	2	4	3	4	4	4	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
ER.13	12	3	3	2	4	3	4	4	4	MR	44.533	-64.499	Environment	7/15/2006	Lunenburg
SR.1.2	12	3	3	2	4	3	4	4	4	MR	44.533	-64.499	Environment	10/31/2006	Lunenburg
S.234	12	3	3	2	4	3	4	4	4	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
S.334	12	3	3	2	4	3	4	4	4	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
S.508	14	9	1	1	7	1	6	1	10	ONth	44.351	-75.990	Environment	8/1/2010	Thousand Islands
32.1	14	9	1	1	7	1	6	1	10	QC	45.329	-73.038	White-tailed deer	11/1/2008	Sainte Brigided Iberville
42	14	9	1	1	7	1	6	1	10	QC	45.178	-73.346	Eastern chipmunk	7/1/2008	Saint Valentin
127	14	9	1	1	7	1	6	1	10	QC	45.064	-73.285	Eastern chipmunk	7/1/2008	Noyan
128	14	9	1	1	7	1	6	1	10	QC	45.064	-73.285	Eastern chipmunk	7/1/2008	Noyan

129	14	9	1	1	7	1	6	1	10	QC	45.064	-73.285	Eastern chipmunk	7/1/2008	Noyan
130	14	9	1	1	7	1	6	1	10	QC	45.064	-73.285	Eastern chipmunk	7/1/2008	Noyan
789	14	9	1	1	7	1	6	1	10	QC	45.545	-73.465	Environment	9/1/2008	Longueuil
BH.2.50	16	2	2	1	2	2	2	2	2	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
F09.29	16	2	2	1	2	2	2	2	2	QC	45.308	-73.011	Environment	6/2/2009	Farnham militaire
F09.42	16	2	2	1	2	2	2	2	2	QC	45.308	-73.011	Environment	6/2/2009	Farnham militaire
CEM.3	19	4	4	3	3	3	3	3	3	MR	44.533	-64.499	Environment	7/15/2006	Lunenburg
S.229	19	4	4	3	3	3	3	3	3	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
S.239	19	4	4	3	3	3	3	3	3	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
S.369	19	4	4	3	3	3	3	3	3	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands
BP-16	29	18	12	1	11	2	15	1	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-3	29	18	12	1	11	2	15	1	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-36	29	18	12	1	11	2	15	1	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-44	29	18	12	1	11	2	15	1	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-45	29	18	12	1	11	2	15	1	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-8	29	18	12	1	11	2	15	1	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
LP-25	29	18	12	1	11	2	15	1	2	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
BP-34	31	20	4	3	3	3	3	3	3	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-6	31	20	4	3	3	3	3	3	3	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-10	32	8	1	1	1	4	16	1	7	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-47	32	8	1	1	1	4	16	1	7	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
25.1	36	10	5	4	6	1	15	1	6	QC	45.058	-73.225	White-tailed deer	11/1/2008	Saint Georges de Clarence ville
BP-1	46	7	6	12	1	2	5	5	5	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-30	46	7	6	12	1	2	5	5	5	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-19	48	8	1	1	14	2	6	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-13	52	12	1	1	7	1	6	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-42	52	12	1	1	7	1	6	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-17	55	23	1	17	20	2	1	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-21	55	23	1	17	20	2	1	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-25	55	23	1	17	20	2	1	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-9	55	23	1	17	20	2	1	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
LP-20	55	23	1	17	20	2	1	1	10	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
BH.2.58	59	6	1	5	1	1	7	1	19	MR	44.533	-64.499	Environment	10/29/2006	Lunenburg
L.12	59	6	1	5	1	1	7	1	19	MR	44.533	-64.499	Deer Mouse	7/15/2006	Lunenburg
S.227	59	6	1	5	1	1	7	1	19	ONth	44.351	-75.990	Environment	6/1/2010	Thousand Islands

102	59	6	1	5	1	1	7	1	19	QC	45.117	-73.212	Eastern chipmunk	6/1/2008	Henryville
264	59	6	1	5	1	1	7	1	19	QC	45.117	-73.212	Environment	5/1/2008	Henryville
F09.51	519	1	1	4	6	1	6	1	1	QC	45.308	-73.011	Environment	6/2/2009	Farnham militaire
LP-76	520	104	1	1	20	2	21	1	2	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
LP-34	521	6	1	15	11	2	20	1	7	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
LP-26	522	8	5	1	6	1	177	1	10	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
LP-13	523	8	1	1	20	1	6	1	10	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
LP-10	524	182	3	17	20	2	4	2	10	ONlp	42.556	-80.197	Environment	11/22/2001	Long Point
BP-46	525	12	1	1	13	1	15	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-43	526	3	3	16	2	3	4	4	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-40	527	3	3	2	2	5	4	4	4	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-11	528	18	2	1	2	1	16	4	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-37	528	18	2	1	2	1	16	4	2	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-35	529	3	1	16	4	1	4	4	4	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-32	530	14	1	5	2	2	1	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-7	530	14	1	5	2	2	1	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-31	531	4	4	16	14	2	4	6	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-27	532	128	1	1	8	1	6	1	10	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-22	533	3	3	1	3	5	4	4	4	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-15	534	137	3	16	4	3	4	4	4	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
BP-2	225	8	2	5	93	2	8	1	84	MB	49.001	-95.233	Environment	4/12/2006	Buffalo Point
L.37	535	9	1	1	6	1	4	1	7	MB	44.533	-64.499	Deer Mouse	7/17/2006	Lunenburg
L.28	536	2	1	1	1	1	7	2	7	MB	44.533	-64.499	Deer Mouse	7/15/2006	Lunenburg
BH.2.5 1	537	6	5	4	6	1	6	1	6	MB	44.533	-64.499	Environment	10/29/2006	Lunenburg
L.38	538	5	1	1	94	1	6	1	7	MB	44.533	-64.499	Deer Mouse	7/17/2006	Lunenburg

Supplementary File 2

Assessment of the possibility that novel MLST STs are due to PCR anomalies of mixed-strain infections of field-collected samples.

Supporting information for Mechai, S, Margos G., Feil EJ, Lindsay LR, Ogden NH (2014) Phylogeographic analysis reveals a complex population structure of *Borrelia burgdorferi* in southern Canada.

Background

It is a hypothesis that novel STs identified in field-collected samples could be spurious assemblages of different alleles for each locus amplified by random chance from different strains in mixed infections. If the hypothesis is true, it would have implications for the use of MLST as a methodology for exploring bacterial phylogeny and strain structure of *B. burgdorferi* and a range of other bacteria. There is circumstantial evidence in the data that this is not the case: i) Mixed infections were found in a proportion of samples using analysis of forward and reverse sequence traces of 9 genes (as has been described for other studies) and these were eliminated from the analysis. ii) The novel STs occurred in a distinct geographic pattern with only one new ST being found in south western Quebec and eastern Ontario even though this region provided the greatest sample size and the bulk of the samples were either adult ticks or engorged nymphs collected from rodents. This geographic pattern is not consistent with novel STs arising from random amplification of mixed infections. iii) For two novel STs, more than one sample carrying the ST was identified and the probability of random assemblages of 9 amplified genes occurring twice in our relatively small sample size would be low. Nevertheless, to test this hypothesis additional empirical are required to support the MLST results.

Material and Methods

We tested the hypothesis that in mixed infections, alleles from the co-infecting strains are not randomly amplified (or at least unpredictably amplified) in each PCR in the MLST, and that consequently the new STs found in Canadian samples were not spurious with different loci coming from different co-infecting strains. When amplifying highly conserved sequences with highly conserved primer binding sites (as all of the MLST loci in this study are), a range of studies (see discussion section) suggest that if two strains are in broadly equal concentrations in the sample this will be detectable in the sequence traces. If the two strains are present in uneven concentrations, that strain present at higher concentration will be preferentially sequenced and that this will occur repeatedly for any conserved locus in the samples.

PCR and sequencing

Two pairs of samples containing different and novel STs were selected from our study and these were re-PCR'd according to the following treatments.

1) The original samples were re-amplified using 2.5µl of extracted DNA solution, without mixing with any other samples, as in the submitted article.
2) Template DNA comprised 1.25µl of extracted DNA solution from each of two samples (that contained different, novel STs according to the original document) in the same PCR, i.e. DNA from the two samples was mixed in equal proportions.
3) The two samples containing different STs were mixed in a ratio of STx : STy of 1 : 10. 1.25µl of extracted DNA solution of the sample containing STy, was mixed with 1.25µl of extracted DNA solution of STx diluted 1/10. This was then used as template for PCR.
4) (3) was repeated but with the ratio of STx : STy being reversed to 10 : 1.

Where possible this was repeated for multiple loci that were different in the pairs of samples and each PCR was repeated in triplicate. The samples for this assay were chosen on the basis of having different novel STs, and having a similar concentration of template DNA assessed on the basis of the cycle threshold for amplification during the PCRs. The pairs of samples chosen were sample IDs BP-7 (containing ST 530) mixed with BP-27 (containing ST 532) and BP-40 (containing ST 708) mixed with BP-43 (containing ST 707). The *clpA, pepX* and *pyrG* loci of the pairing BP-7 and BP-27, and the *nifS* locus of the pairing BP-40 and BP-43 were amplified and sequenced as described in the original document. The two samples containing the same, novel ST in the main manuscript (samples BP-11 and BP-37 containing ST 528) were also subject to re-PCR and PCR after mixing of DNA as described above but the targets were *clpA* and *uvrA*. The expected outcomes were as follows: i) re-PCR of unmixed samples would produce the same allele of each locus as in the original amplification and sequencing; ii) mixing of sample DNA in equal quantities would produce traces clearly showing two different amplified nucleotides at the positions where the two different alleles differed; and iii) mixing of sample DNA in unequal quantities would produce traces showing a higher peak for the nucleotide amplified from the higher concentration allele with, in most cases, a lower peak for the nucleotide amplified from the low concentration allele, at the positions where the two different alleles differed. We also expected that in the same mixtures of samples expected

outcomes ii) and iii) would be consistent for all loci, that findings would be consistent for all replicates, and that when the samples mixed together were of the same ST, only one allele would be detectable in sequence traces.

Statistical analysis

The null hypothesis was that the expected outcomes for PCRs of mixed samples as described above would occur no more frequently than by random chance. The most conservative estimate of probability of the outcome of amplification and sequencing would be that based on there being only two STs in the mixed samples. This is because were amplification of different alleles to be random and not dependent on ST DNA concentrations, the more STs in the samples the more likely an unexpected outcome would be. Assuming that there are two STs only in each mixture, the outcome under the null hypothesis would be that approximately 50% of PCRs yield unexpected outcomes. If the proportion of outcomes that are expected is significantly greater than the lower 95% confidence interval for A/B = 0.5 (where A is the number of unexpected results and B is the total number of PCRs of mixed samples), then we considered that expected outcomes occurred more frequently than by chance and that the null hypothesis was rejected. The level of significance was $P < 0.05$.

Results

The results are shown in Table 1 and sequence traces for one example replicate for each unmixed and mixed sample is shown in Supplementary File 3. All re-PCR of samples without mixing yielded the same allele as in the original manuscript. All 36 reactions comprising mixtures of template DNA from samples carrying different STs produced sequences as expected: mixing of sample DNA in equal quantities produced traces showing two different amplified nucleotides at the positions where the two different alleles differed; mixing of sample DNA in unequal quantities produced traces showing two different amplified nucleotides at the positions where the two different alleles differed but the dominant allele having a higher peak on the sequence trace. These findings were consistent for all three loci PCR'd and sequenced for the mixture of samples BP-7 and BP-27. Under the null hypothesis the proportion of unexpected results would be approximately 50% (18/36) and the number of unexpected results at the lower exact binomial 95% confidence interval (i.e. the minimum

number of unexpected results that is not significantly different from 18) is 12/36. The observed number of unexpected results was 0/36, which was significantly different from 12/36 ($\chi^2 = 14.4$, df=1, $P < 0.001$).

In addition, i) *clpA*, *pepX* and *pyrG* loci from BP-7 and BP-27 and the *nifS* locus from BP-40 and BP-43 were re-PCR'd in triplicate without mixing and the 24 PCRs yielded the same original allele each time, and ii) two loci from samples of the same novel ST were amplified with and without mixing and again yielded the same original allele each time of 30 PCRs. Therefore, on these essentially 'unmixed' samples the 54/54 PCRs yielded the expected result. If these samples did contain mixed infections that resulted in random amplification of different alleles, then this result would be the equivalent of tossing a coin 54 times and obtaining the same result each time, which has a probability of 5.5×10^{-17} using the formula $(1/2)^{54}$).

Discussion

In this study, using samples collected in the field and carrying novel STs, none of the PCRs produced unexpected results: multiple re-PCRs of the loci repeatedly produced the same alleles as in the original PCR. Mixing of samples resulted in different alleles being amplified consistently according to which was the dominant in terms of the proportion of template DNA in the mix, and samples mixed in equal proportions produced mixed alleles easily detectable in the sequence traces. These findings were consistent for two different mixtures of samples, and for one pair of samples were consistent at three loci. No other alleles other than those originally detected were amplified when two independent samples that carried the same novel ST were mixed. Together these findings support the conclusions that i) alleles amplified in the original PCRs that were used to assign samples to particular STs are consistently amplified and do not constitute randomly-amplified alleles in mixed infections; ii) analysis of sequences in the traces is a reliable method of identifying mixed infections as long as the mixed strains are in broadly similar concentrations; iii) where mixed-strain infections occur but one strain dominates in terms of copy numbers, the dominant strain will produce the clearest peaks in sequencing traces and this is consistent for multiple loci of the strains when amplified from the same samples.

There are particular characteristics of *B. burgdorferi*, the samples from which DNA was obtained, and of the loci amplified in the MLST, that may make the findings of this study

more likely compared to MLST applied to bacteria in other environmental samples. The occurrence of mixed strains of *B. burgdorferi* in field collected samples is limited by the occurrence of this bacterium in relatively low copy numbers (a few thousands of bacteria or less) in many of its niches in rodents and ticks during the transmission cycle (1-5), the occurrence of innate immunity, acquired immunity and cross-strain interactions in reservoir hosts as well as a range of population bottlenecks in ticks during the transmission cycle limit the range and frequency of transmissible mixed infections (2, 5, 6, 7). Furthermore host-seeking nymphal ticks will have fed on only one host and while engorged nymphal and host-seeking adult ticks will have fed on two hosts prior to DNA extraction, bottlenecks during the transmission cycle appear to limit increases in strain diversity (5, 8). The loci used in the MLST are highly conserved housekeeping genes and the alleles differ by substitutions of a very small number of nucleotides (usually less than 5) within a genospecies. Therefore the sequences are highly similar and do not vary in length reducing the likelihood that one allele would be preferentially amplified over another when two or more co-occur in mixed infections. In this respect the loci in the MLST are similar to those in a range of studies in which Sanger sequencing is explicitly used as a tool to identify the occurrence of mixed-allele samples (9-11) but the principles may have wider relevance in microbiology (12). We cannot be certain that every novel ST identified in every field-collected sample, after elimination of samples that show evidence of mixed infections on the basis of evaluating traces from Sanger sequencing, are not artefactual and due to amplification of loci from different strains of the mixed infection. However this study suggests that if that does occur, it would be a very low probability event.

References

1. **Burkot TR, Piesman J, Wirtz RA.** 1994. Quantitation of the *Borrelia burgdorferi* outer surface protein A in *Ixodes scapularis*: fluctuations during the tick life cycle, doubling times, and loss while feeding. J. Infect. Dis. **170**:883-889.

2. **Liang FT, Brown EL, Wang T, Iozzo RV, Fikrig E.** 2004. Protective niche for *Borrelia burgdorferi* to evade humoral immunity. Am. J. Pathol. **165**:977-985.

3. **Soares CA, Zeidner NS, Beard CB, Dolan MC, Dietrich G, Piesman J.** 2006. Kinetics of *Borrelia burgdorferi* infection in larvae of refractory and competent tick vectors. J. Med. Entomol. **43**:61-67.

4. **Hojgaard A, Eisen RJ, Piesman J.** 2008. Transmission dynamics of *Borrelia burgdorferi* s.s. during the key third day of feeding by nymphal *Ixodes scapularis* (Acari: Ixodidae). J Med Entomol. **45**:732-736.

5. **Rego RO, Bestor A, Stefka J, Rosa PA.** 2014. Population bottlenecks during the infectious cycle of the Lyme disease spirochete *Borrelia burgdorferi*. PLoS One. **9**:e101009.

6. **Derdáková M, Dudiòák V, Brei B, Brownstein JS, Schwartz I, Fish D.** 2004. Interaction and transmission of two *Borrelia burgdorferi* sensu stricto strains in a tick-rodent maintenance system. Appl. Environ. Microbiol. **70**:6783-6788.

7. **Troy EB, Lin T, Gao L, Lazinski DW, Camilli A, Norris SJ, Hu LT.** 2013. Understanding barriers to *Borrelia burgdorferi* dissemination during infection using massively parallel sequencing. Infect. Immun. **81**:2347-23578. **Rauter C, Hartung T.** 2005. Prevalence of *Borrelia burgdorferi* sensu lato genospecies in *Ixodes ricinus* ticks in Europe: a meta-analysis. Appl. Environ. Microbiol. **71**:7203–7216.

9. **Li J, Wang L, Mamon H, Kulke MH, Berbeco R, Makrigiorgos GM.** 2008. Replacing PCR with COLD-PCR enriches variant DNA sequences and redefines the sensitivity of genetic testing. Nat. Med. **14**:579–584.

10. **Milbury CA, Li J, Makrigiorgos GM.** 2010. Ice-COLD-PCR enables rapid amplification and robust enrichment for low-abundance unknown DNA mutations. Nucleic Acids Res. **39**:e2.

11. **Tsiatis AC, Norris-Kirby A, Rich RG, Hafez MJ, Gocke CD, Eshleman JR, Murphy KM.** 2010. Comparison of Sanger sequencing, pyrosequencing, and melting curve analysis for the detection of KRAS mutations: diagnostic and clinical implications. J. Mol. Diagn. **12**:425-432.12. **Eyre DW, Cule ML, Griffiths D, Crook DW, Peto TEA, Walker AS, Wilson DJ.** 2013. Detection of mixed infection from bacterial whole genome sequence data allows assessment of its role in *Clostridium difficile* transmission. PLOS Comput. Biol. **9**:e1003059.

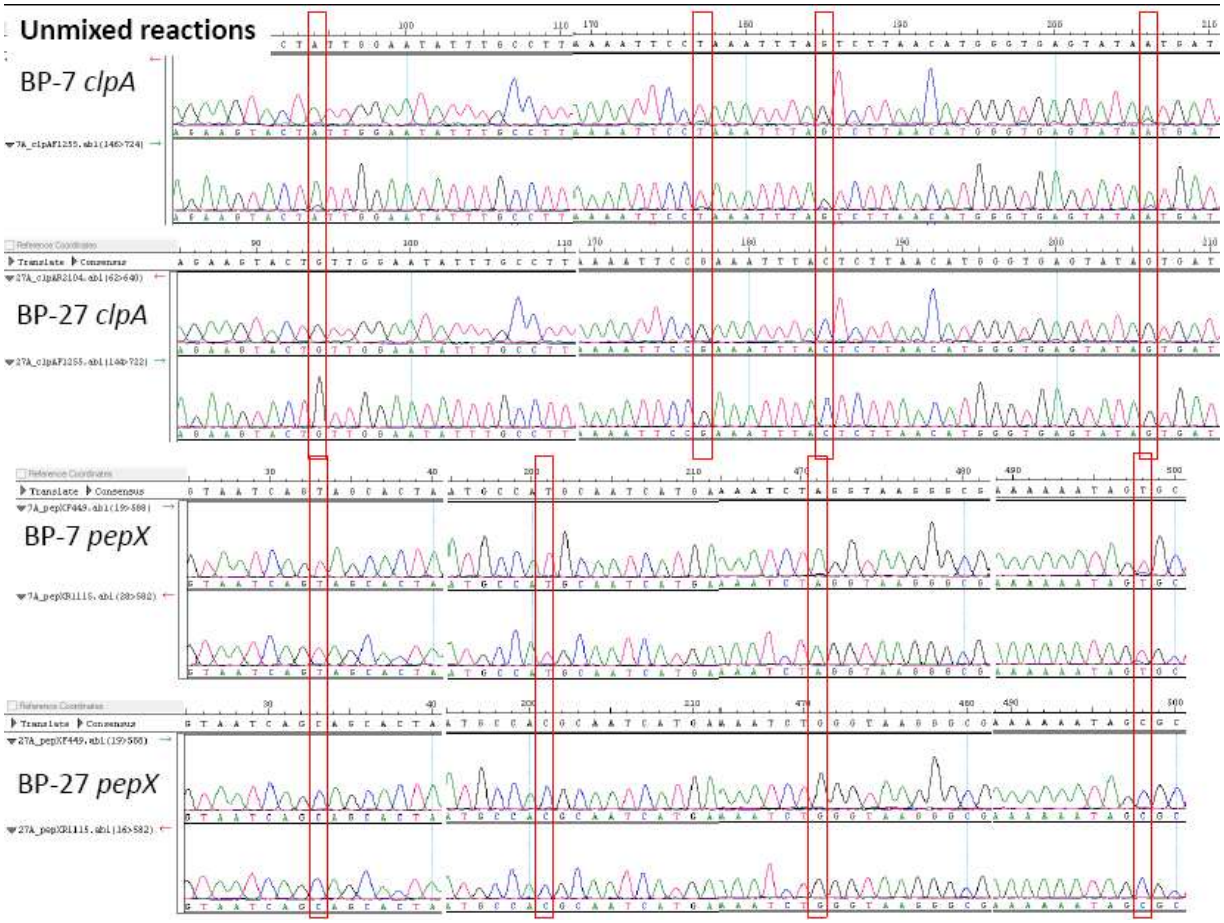
Table 1. Pattern of sample mixing and PCR-product sequencing results. NA = not applicable.

Sample IDs	Ratio of DNA in mixed samples	Replicate	Dominant allele in sequence trace				
			clpA	*pyrG*	*uvrA*	*pepX*	*nifS*
BP-7	NA	A	14	2		2	
BP-7	NA	B	14	2		2	
BP-7	NA	C	14	2		2	
BP-27	NA	A	128	1		8	
BP-27	NA	B	128	1		8	
BP-27	NA	C	128	1		8	
BP-7 + BP-27	9:1	A	14	2		2	
BP-7 + BP-27	9:1	B	14	2		2	
BP-7 + BP-27	9:1	C	14	2		2	
BP-7 + BP-27	1:9	A	128	1		8	
BP-7 + BP-27	1:9	B	128	1		8	
BP-7 + BP-27	1:9	C	128	1		8	
BP-7 + BP-27	1:1	A	Mixed	Mixed		Mixed	
BP-7 + BP-27	1:1	B	Mixed	Mixed		Mixed	
BP-7 + BP-27	1:1	C	Mixed	Mixed		Mixed	
BP-40	NA	A					2
BP-40	NA	B					2
BP-40	NA	C					2
BP-43	NA	A					16
BP-43	NA	B					16
BP-43	NA	C					16
BP-43 + BP-40	9:1	A					16
BP-43 + BP-40	9:1	B					16
BP-43 + BP-40	9:1	C					16
BP-43 + BP-40	1:9	A					2
BP-43 + BP-40	1:9	B					2
BP-43 + BP-40	1:9	C					2
BP-43 + BP-40	1:1	A					Mixed
BP-43 + BP-40	1:1	B					Mixed
BP-43 + BP-40	1:1	C					Mixed
BP-11	NA	A	18		2		

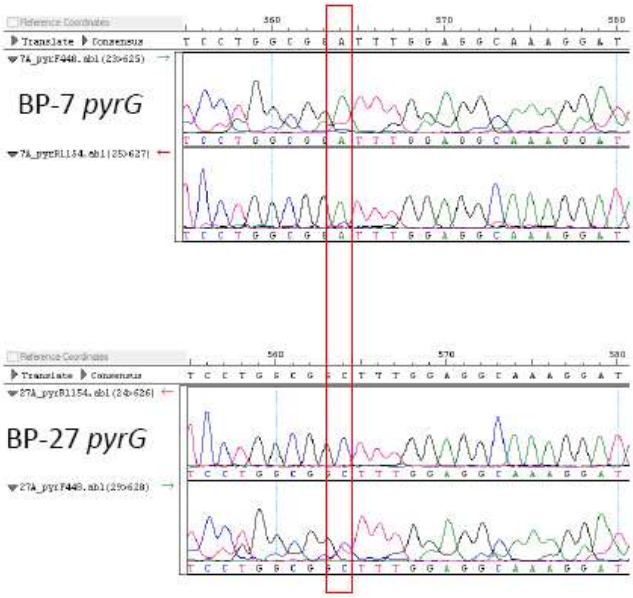
BP-11	NA	B	18	2
BP-11	NA	C	18	2
BP-37	NA	A	18	2
BP-37	NA	B	18	2
BP-37	NA	C	18	2
BP-11 + BP-37	9:1	A	18	2
BP-11 + BP-37	9:1	B	18	2
BP-11 + BP-37	9:1	C	18	2
BP-11 + BP-37	1:9	A	18	2
BP-11 + BP-37	1:9	B	18	2
BP-11 + BP-37	1:9	C	18	2
BP-11 + BP-37	1:1	A	18	2
BP-11 + BP-37	1:1	B	18	2
BP-11 + BP-37	1:1	C	18	2

Supplementary File 3:

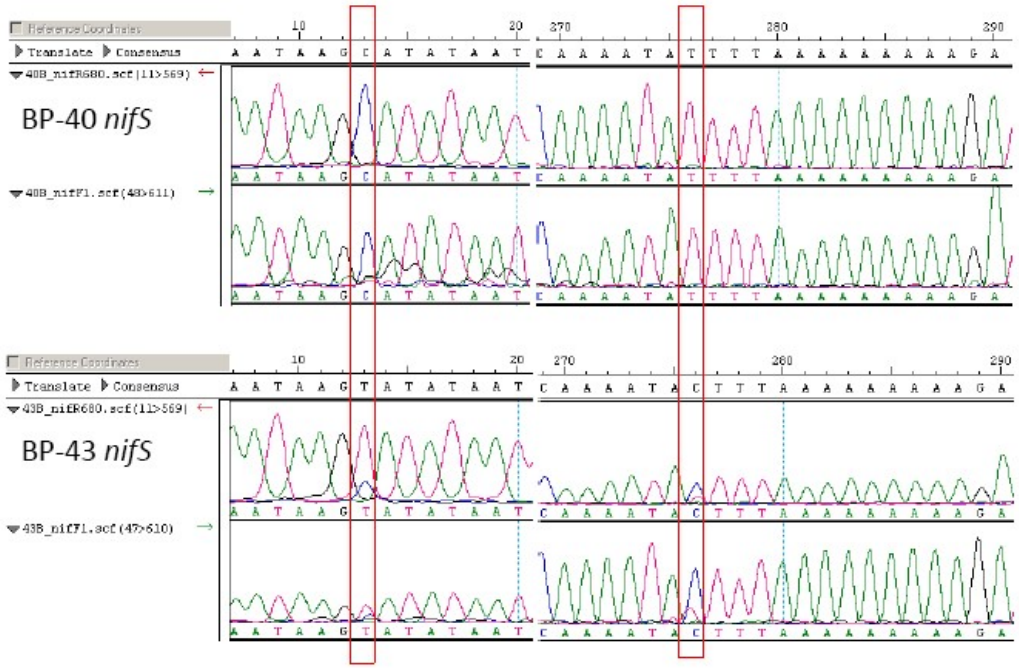
Supplementary Figure 1. Sequence traces of loci amplified from the different alleles of unmixed and mixed samples containing novel STs. Where samples were mixed the relative proportions of template DNA from each are shown. Nucleotides that differ between pairs of mixed samples are identified by red boxes.



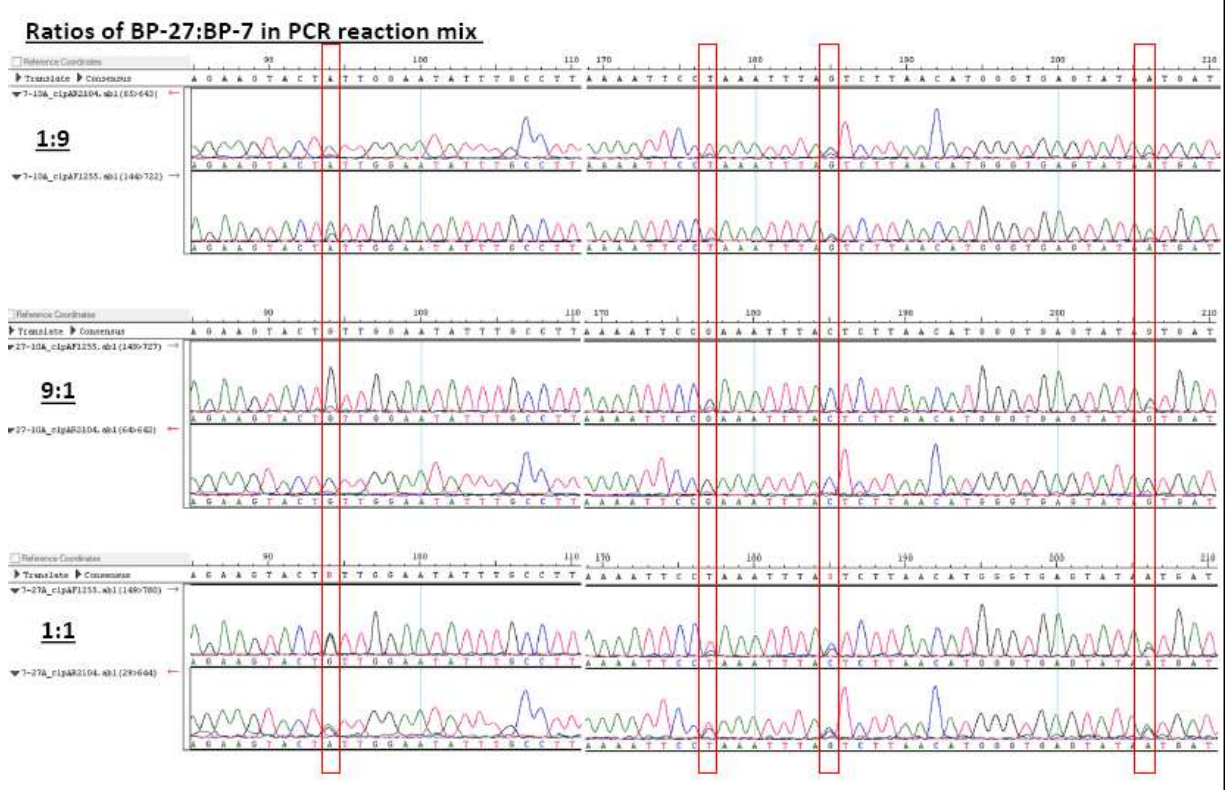
Unmixed reactions



Unmixed reactions

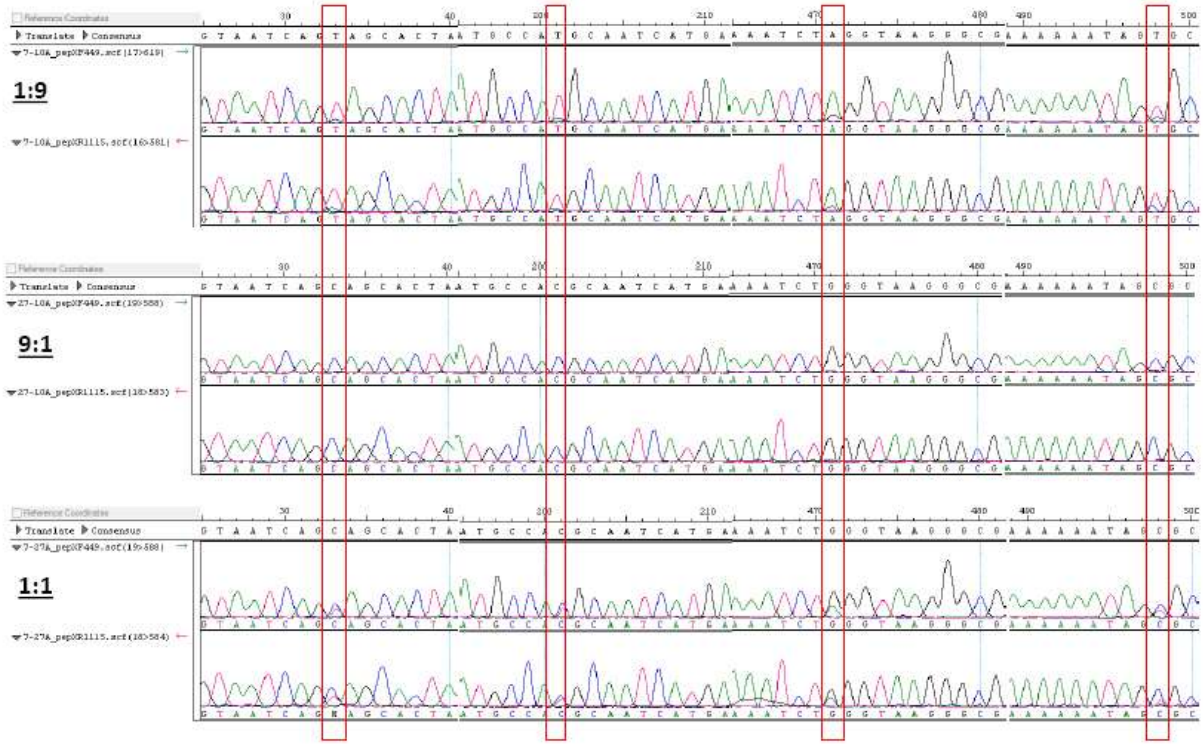


Mixed reactions of BP-7 and BP-27: *clpA* locus



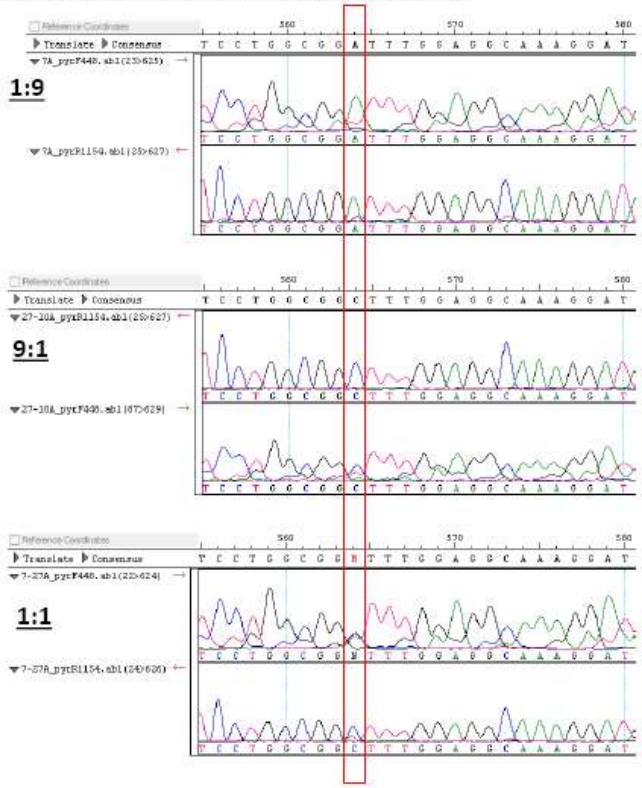
Mixed reactions of BP-7 and BP-27: *pepX* locus

Ratios of BP-27:BP7 in PCR reaction mix



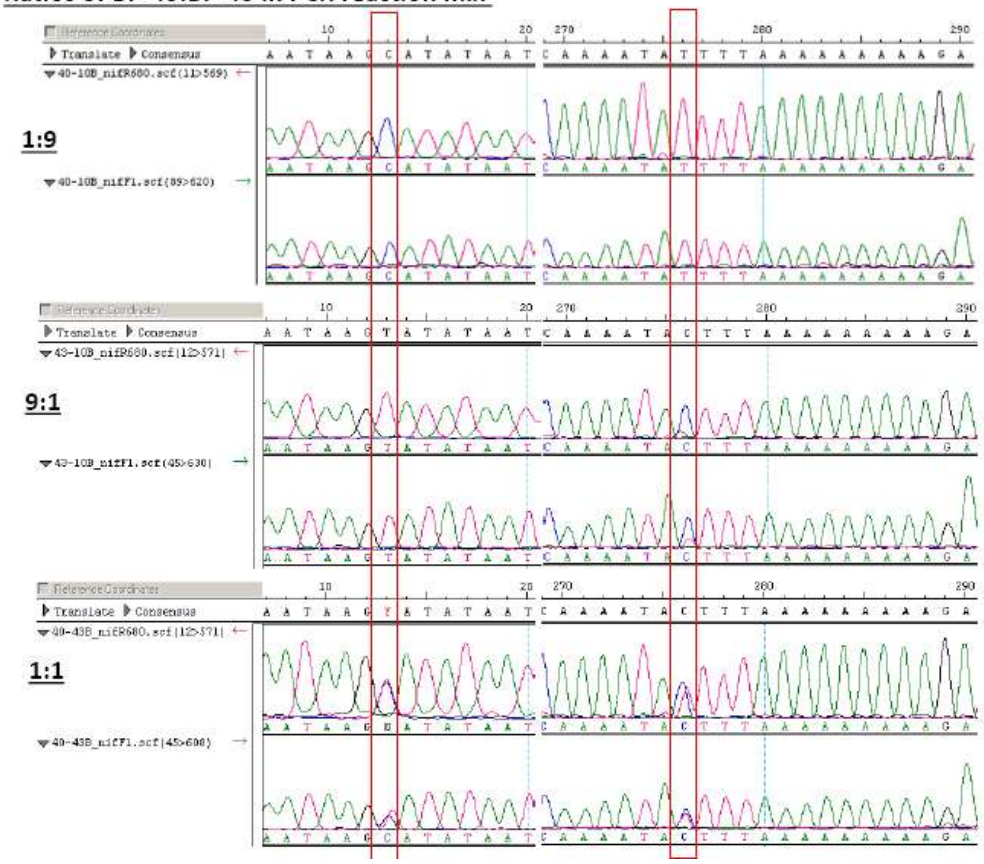
Mixed reactions of BP-7 and BP-27: *pyrG* locus

Ratios of BP-27:BP-7 in PCR reaction mix



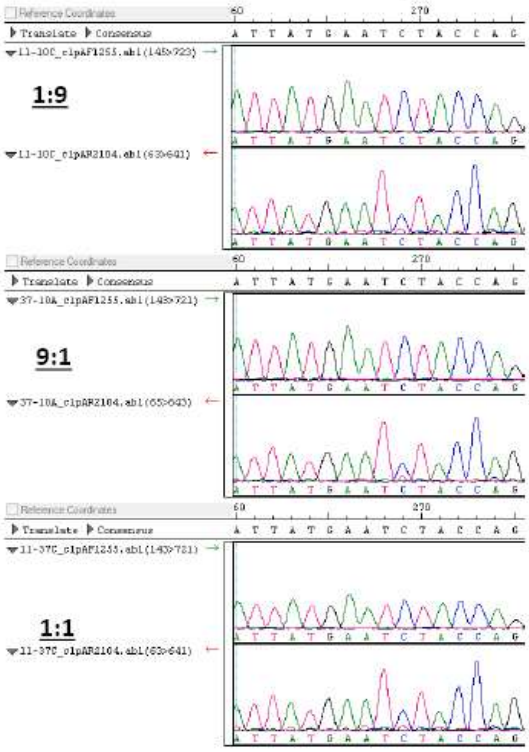
Mixed reactions of BP-43 and BP-40: *nifS* locus

Ratios of BP-43:BP-40 in PCR reaction mix



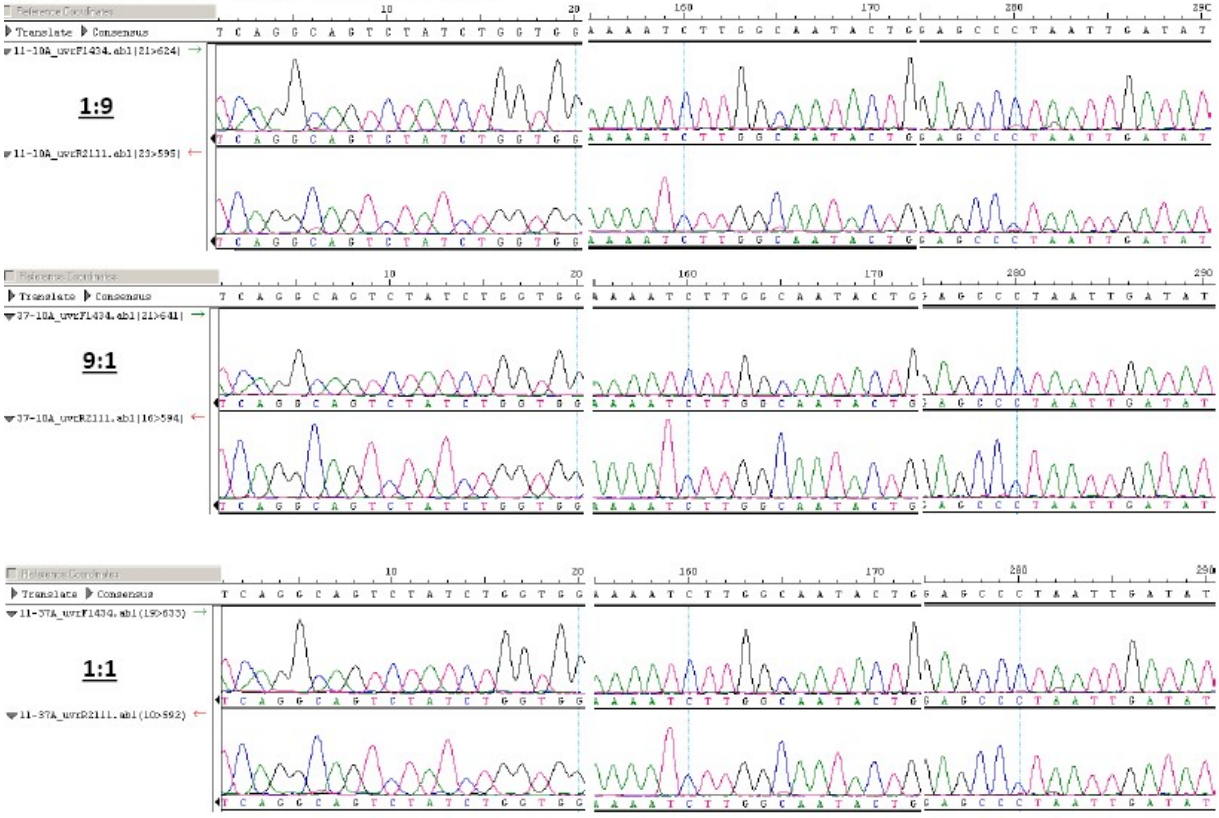
Mixed reactions of BP-11 and BP-37: *clpA* locus

Ratios of BP-11:BP-37 in PCR reaction mix



Mixed reactions of BP-11 and BP-37: *uvrA* locus

Ratios of BP-37:BP-11 in PCR reaction mix



IV. Chapitre II : *Investigation des facteurs déterminants l'occurrence des souches de B. burgdorferi au Canada*

Evidence for Host-Genotype associations of *Borrelia burgdorferi* sensu stricto

Samir Mechai,[1] Gabriele Margos,[2,3] Edward J. Feil,[4] Nicole Barairo,[5] L. Robbin Lindsay,[5] Pascal Michel,[1,6] and Nicholas H. Ogden[1,6,*]

[1]Groupe de Recherche en Épidémiologie des Zoonoses et Santé Publique, Faculté de Médecine Vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, Canada
[2]National Reference Centre for *Borrelia*, Oberschleissheim, Germany
[3]Bavarian Health and Food Safety Authority, Oberschleissheim, Germany
[4]The Milner Centre for Evolution, Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom
[5]National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada
[6]National Microbiology Laboratory, Public Health Agency of Canada, Saint-Hyacinthe, Québec, Canada

Citation: Mechai S, Margos G, Feil EJ, Barairo N, Lindsay LR, Michel P, et al. (2016) Evidence for Host-Genotype Associations of *Borrelia burgdorferi* Sensu Stricto. PLoS ONE 11(2): e0149345. doi:10.1371/journal.pone.0149345

Brian Stevenson, Editor
University of Kentucky College of Medicine, UNITED STATES
Received: November 5, 2015
Accepted: January 29, 2016
Published: February 22, 2016. doi: 10.1371/journal.pone.0149345
Competing Interests: The authors have declared that no competing interests exist.
Conceived and designed the experiments: NHO SM GM EF PM. Performed the experiments: NB LRL. Analyzed the data: SM NHO EF GM PM. Contributed reagents/materials/analysis tools: EF GM. Wrote the paper: SM GM EF LRL PM NHO.
* E-mail: nicholas.ogden@phac-aspc.gc.ca

Abstract

Different genotypes of the agent of Lyme disease in North America, *Borrelia burgdorferi* sensu stricto, show varying degrees of pathogenicity in humans. This variation in pathogenicity correlates with phylogeny and we have hypothesized that the different phylogenetic lineages in North America reflect adaptation to different host species. In this

study, evidence for host species associations of *B. burgdorferi* genotypes was investigated using 41 *B. burgdorferi*-positive samples from five mammal species and 50 samples from host-seeking ticks collected during the course of field studies in four regions of Canada: Manitoba, northwestern Ontario, Quebec, and the Maritimes. The *B. burgdorferi* genotypes in the samples were characterized using three established molecular markers (multi-locus sequence typing [MLST], 16S-23S *rrs-rrlA* intergenic spacer, and outer surface protein C sequence [*ospC*] major groups). Correspondence analysis and generalized linear mixed effect models revealed significant associations between *B. burgdorferi* genotypes and host species (in particular chipmunks, and white-footed mice and deer mice), supporting the hypotheses that host adaptation contributes to the phylogenetic structure and possibly the observed variation in pathogenicity in humans.

Introduction

In North America, *Borrelia burgdorferi* sensu stricto (hereafter termed *B. burgdorferi* for simplicity) is a member of the bacterial genospecies complex *B. burgdorferi* sensu lato (s.l.) that is associated with Lyme disease [1]. In Eurasia, five genospecies of the *B. burgdorferi* s.l. complex are associated with Lyme disease [1]: *B. burgdorferi*, *B. afzelii*, *B. garinii*, *B. bavariensis* and *B. spielmanii* and the two main tick vectors are *Ixodes ricinus* (in Europe) and *I. persulcatus* (in Asia) [2, 3]). In North America, *B. burgdorferi* is mostly transmitted by two tick species: *I. scapularis* in the regions encompassing northeastern USA and southeastern Canada, and the upper Midwest USA and south central Canada, and *I. pacificus* in the western coastal states of the USA and in British Columbia, Canada.

In Eurasia, the different *B. burgdorferi* s.l. genospecies are associated with different types of clinical disease [4]. Arthritis is associated with *B. burgdorferi* infection; neuroborreliosis with *B. garinii* and *B. bavariensis* infection, and chronic dermatological manifestations with *B. afzelii* [5–7]. Most of the clinical features seen in Europe are also seen in North America, and these include those of early Lyme disease (Erythema migrans: EM), early disseminated Lyme disease (neuroborreliosis including facial palsy, meningitis and peripheral radiculoneuropathy, and atrioventricular block) and late disseminated Lyme disease (including Lyme arthritis) [2,

8, 9]. In North America there is evidence that different genotypes of *B. burgdorferi* show different levels of pathogenicity in humans, specifically whether or not the bacterium disseminates systemically from the early phase infection in the skin where the infective tick bit the patient [10–14]. In Europe, *B. burgdorferi* s.l. genospecies are frequently specialized for transmission by different host species [15]: *B. afzelii* and *B. bavariensis* are rodent host specialists [15, 16], *B. garinii* is a bird specialist [8] and *B. lusitaniae* may be a lizard specialist [17]. In North America, *B. burgdorferi* is considered a host generalist [4, 18], although more stable suitable environments associated with expanding woodland habitats, increased abundance of tick vectors and reservoir hosts [19] are thought to be creating conditions favourable for adaptive radiation and multiple niche polymorphism [4]. Most parasites show some degree of host preference [20–22], which is a critical pre-adaptation for host specialization if conditions for transmission are suitable [23], as they may increasingly be for *B. burgdorferi* in North America. There is some evidence of host associations for *B. burgdorferi* in the form of unequal frequencies of *B. burgdorferi* genotypes in samples collected in the field from different sources [24–26], and differential infection and transmission efficiency among different host-genotype pairings [25, 27]. Such associations are of public health interest as they may be linked to the capacity of the different genotypes to show different pathogenicity in humans and varying capacity to stimulate antibodies detectable in current serological tests, while the existence of host associations may allow prediction of regions and habitats where different genotypes are more likely to occur [28].

The clade structure of the *B. burgdorferi* phylogenetic tree obtained using concatenated housekeeping genes of a multi-locus sequence typing (MLST) method does not seem to be based on geographic isolation of genotypes [29, 30]. It has been hypothesized that the clades were associated with introductions and/or population expansions after bottlenecks possibly associated with glacial-interglacial periods [31], although ecological isolation driven by host species associations may also explain the origin and maintenance of discrete clusters [28]. Small and medium-sized vertebrates, particularly rodents, are key requirements for *B. burgdorferi* transmission cycles in northern North America as these species are frequently competent reservoirs of *B. burgdorferi* and important hosts for immature ticks. Adult ticks feed preferentially on larger mammals, mostly reservoir-incompetent deer [32]. The near

absence of *B. burgdorferi* from *I. scapularis* ticks in southeastern USA is thought to be associated in part with the high proportion of immature ticks in this region that feed on reservoir-incompetent lizards and the low proportion that feed on reservoir-competent rodents [33].

Lyme disease is currently emerging in central and southeastern Canada associated with the northward expansion of the geographic range of *I. scapularis*, which is possibly associated with climate change [34, 35]. Range expansion of both *I. scapularis* and *B. burgdorferi* is likely being facilitated by dispersal of ticks and bacteria by both migratory birds and terrestrial hosts [36, 37]. It is also possible that refugial populations of *B. burgdorferi* are maintained by nidicolous ticks [38] but these have yet to be found in Canada.

A complex geographic pattern of genotypes has been found in emerging Lyme disease risk areas in south central and southeastern Canada [30], and in this study we explore possible association of *B. burgdorferi* genotypes with host species using samples collected in these regions.

Material and Methods

Samples used in the study

The samples used in this study comprise DNA of *B. burgdorferi* extracted from host-seeking *I. scapularis* ticks, engorged ticks from captured rodent hosts and from *B. burgdorferi*-positive tissues from rodent hosts. If there were multiple samples from the same rodent (which could be either an engorged immature tick or a tissue sample), only one sample per rodent host was randomly selected for inclusion in the study. All samples used in host association analyses in this study were collected during field studies in 41 woodland locations in south central and southeastern Canada from 2006 to 2013. In these studies rodents were captured using Sherman traps, examined for ticks under anaesthesia, and subsequently euthanized by cervical dislocation under anaesthesia. Feeding ticks and rodent tissues were collected and transferred

to the laboratory for testing for *B. burgdorferi* as previously described [39, 40]. Host-seeking ticks were collected by drag sampling and also transferred to the laboratory for testing for *B. burgdorferi* [40]. Rodents were trapped on private properties that did not require specific permissions and did not involve the capture of endangered or protected species. All rodents were captured and dispatched using protocols approved by animal care committees of either the Canadian Science Centre for Human and Animal Health or Université de Montréal, and with relevant scientific collection permits for the locations where the work was conducted (Manitoba Conversation, Ontario Ministry of Natural Resources, Ministère des Ressources naturelles et de la Faune du Québec and the Nova Scotia Department of Natural Resources).

For the host association analyses in this study there was a total of 91 *B. burgdorferi*-positive samples from the sites in Canada including 41 samples from 5 rodent species namely: deer mouse (*Peromyscus maniculatus* Wagner, 1845, n = 2), eastern chipmunk (*Tamias striatus* Linnaeus, 1758, n = 16), red squirrel (*Tamiasciurus hudsonicus* Erxleben, 1777, n = 2), red backed vole (*Myodes gapperi* Vigors, 1830, n = 1) and white-footed mouse (*Peromyscus leucopus* Rafinesque, 1818, n = 20), as well as 50 questing ticks collected contemporaneously with the samples from rodents (Table 1). DNA was extracted from all samples (questing ticks, engorged ticks and host tissue samples) and screened for *B. burgdorferi* infection by polymerase chain reaction (PCR) as previously described [30, 36]. Sampling effort was not consistent at each site visit, and this study therefore consists of a convenience sample. However this is accounted for in analyses as described in the following sections.

Table 1: Data on ST frequencies among hosts/sources.
The number of *B. burgdorferi* MLST sequence types among 5 host species used in this study: *Peromyscus maniculatus* (DM: Deer mouse), *P. leucopus* (WFM: White-footed mouse), *Tamias striatus* (ECH: Eastern chipmunk), *Tamiasciurus hudsonicus* (RS: Red squirrel) and Myodes gapperi (RBV: Redbacked vole), and host-seeking ticks (Ticks) sampled in four different regions from Southern Canada: Manitoba (MB), Ontario at Rainy River (ONRv), Quebec (QC) and the Maritimes (MR).

ST	DM	ECH	RBV	RS	WFM	Ticks	MB	ONRv	QC	MR
1	0	0	1	0	7	9	0	0	9	8
3	0	1	0	1	4	14	0	0	11	9
4	0	0	0	0	1	1	0	0	1	1
8	0	1	0	0	0	1	0	0	2	0
9	0	0	0	0	3	3	0	0	2	4
12	0	1	0	0	0	7	0	1	4	3
14	0	3	0	0	0	0	0	0	2	1
16	0	0	0	0	0	4	0	0	3	1
19	0	1	0	0	0	2	0	0	1	2
29	0	1	0	0	0	0	0	1	0	0
32	1	0	0	0	0	0	0	1	0	0
36	0	0	0	0	1	1	0	0	1	1
46	0	0	0	0	0	1	1	0	0	0
59	0	1	0	0	1	2	0	0	2	2
222	0	2	0	0	0	0	0	2	0	0
225	0	0	0	1	0	0	0	1	0	0
228	0	1	0	0	0	0	0	1	0	0
234	0	1	0	0	0	0	0	1	0	0
300	0	1	0	0	0	0	0	1	0	0
302	1	0	0	0	0	0	1	0	0	0
315	0	0	0	0	0	1	0	0	1	0
519	0	0	0	0	0	1	0	0	1	0
532	0	1	0	0	0	0	0	1	0	0
535	0	0	0	0	1	0	0	0	0	1
536	0	0	0	0	1	0	0	0	0	1
537	0	0	0	0	0	1	0	0	0	1
538	0	0	0	0	1	0	0	0	0	1
641	0	1	0	0	0	0	0	1	0	0
643	0	0	0	0	0	1	0	0	1	0
644	0	0	0	0	0	1	0	0	1	0
Total	2	16	1	2	20	50	2	11	42	36

Locations of sites where ticks and/or rodents were collected include those presented in [26] as well as more recently visited sites in northwestern Ontario. To facilitate analyses, study sites that were in close proximity were grouped into six geographic regions as follows: (1)

Manitoba (MB; 8 sites), (2) Ontario Rainy River (ONRv; 4 sites), (3) Ontario Long Point (ONLp; 1 site); (4) Ontario East (ONEst; 7 sites), (5) Quebec (QC; 16 sites), and (6) the Maritimes (MR; 14 sites). All samples from these regions were used in phylogenetic analyses, however only questing ticks, without contemporaneously collected rodent host samples, were available from ONLp and ONEst. Data from only four regions (MB, ONRv, QC and MR), where rodent samples and questing ticks were collected contemporaneously, were therefore used in statistical analyses, and the locations of the sites in these regions are shown in Fig 1. The full range of sites in the US and Canada where samples have been collected for phylogenetic analysis (excluding those from ONRv, which are the most recently sampled sites) is shown in Fig 1 of reference [30]. The regions of MB and ONRv combined, and QC and MR comprise regions of emergence of *B. burgdorferi* in Canada and we have no reason to believe that the sites from which rodent samples were collected in this study were in any way outliers compared to other sites in these regions in terms of rodent host species (with the caveat that deer mice predominate over white-footed mice in the more western regions and vice versa in the eastern regions) and *B. burgdorferi* genotypes.

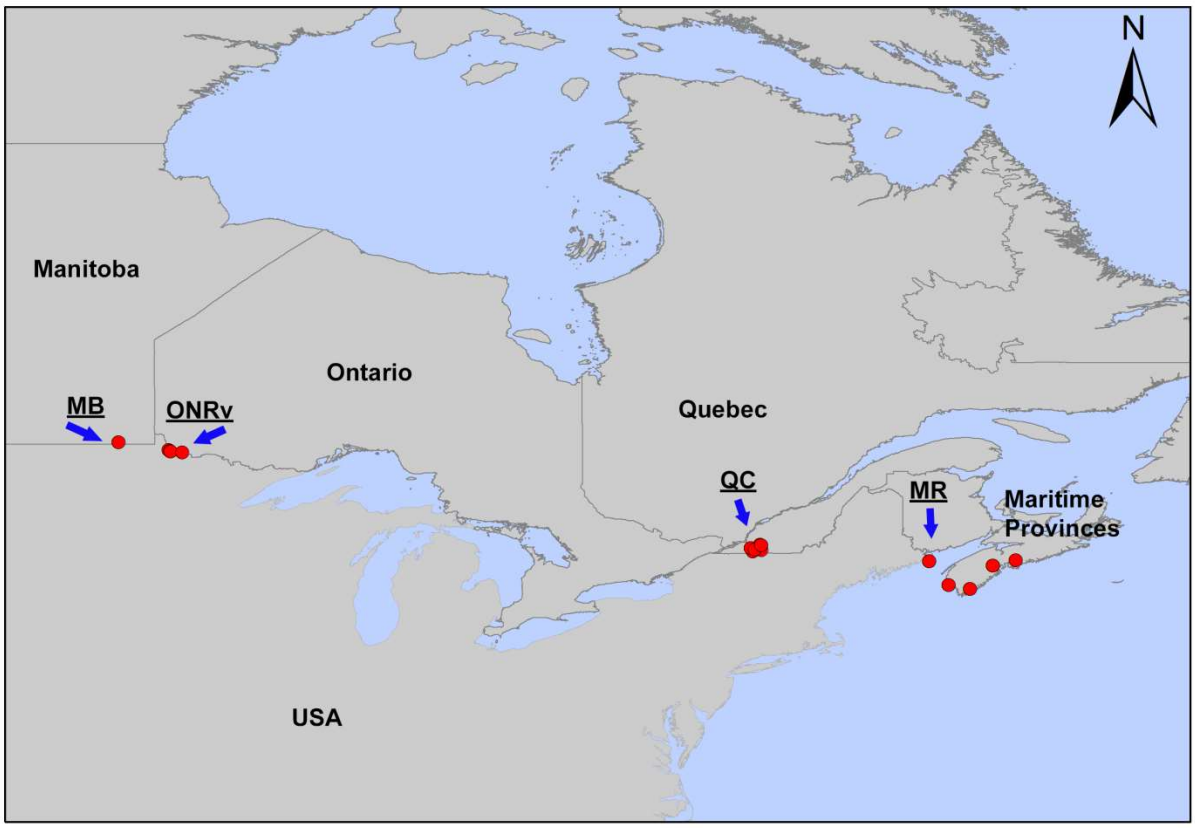


Figure 1: Geographic distribution of the sampling sites in Canada.
A map of northern North America showing the names of the main Canadian provinces. Red dots indicate the locations of sample sites in each of the four regions where samples from animal hosts were available which are indicated by the abbreviations used in the text (MB = Manitoba, ONRv = Rainy River Ontario, QC = southern Quebec and MR = the Maritime provinces of Nova Scotia and New Brunswick).

Genotyping B. burgdorferi

For this study, we focused on genotyping by MLST using eight chromosomal housekeeping genes as previously described [29, 41]. However, we also amplified, sequenced and analysed 16S-23S rrs-rrlA intergenic spacer (IGS) sequences, as well as the outer surface protein C gene (ospC), which have both been used in genotyping of *B. burgdorferi* [42, 43]. Any samples that showed evidence of mixed-genotype infections on examination of the housekeeping genes and IGS sequences were excluded [30]. Of the 91 samples analysed for

the first time in this study, all had MLST data, 80 had IGS data and 69 had ospC data. The difference in the number of samples was due to the lack of available DNA for some samples.

Genotyping using the MLST scheme was conducted as previously described [30, 41]. Fragments of the eight housekeeping genes (clpA, clpX, nifS, pepX, pyrG, recG, rplB and uvrA) were amplified, and the resulting sequences were assigned to existing or new allele numbers (for novel sequences) using the MLST database (http://www.pubmlst.org). The allele combination for each sample was assigned to an existing or to a new sequence type (ST) number (for genotypes with novel alleles or novel allele combinations). All data from this study are available at http://www.pubmlst.org. Clusters of related STs were then identified as follows. MLST STs were 'classified' into clonal complexes obtained using eBURST V.3 [44] to reconstruct relationships between *B. burgdorferi* STs identified in this study. This clustering method allows clonal complexes to be constructed using different criteria for the relatedness of the STs: single locus variants and double locus variants (SLV and DLV). Each sample was assigned membership to clonal complexes defined using criteria of both SLV and DLV. The confidence in the relationship of the member of each clonal complex was computed by the spanning edge betweenness (SEB) corresponding to the percentage of the equivalent minimum spanning trees (MSTs) between STs of the same clonal complex (i.e. the optimal edge selected by the goeBURST algorithm is the most frequently reproduced edge in the MST forest of the clonal complex), which is expressed as a bootstrap value [45]. An unrooted Bayesian phylogenetic tree of the aligned STs without outgroups was constructed using MrBayes v3.2.1 [46] to support the SEB values, which allowed the posterior probability of each corresponding clonal complex at SLV and DLV to be deduced, and for CCs to be visualised alongside the different clades of *B. burgdorferi*. A rooted Bayesian phylogenetic tree of the aligned STs was also generated using MrBayes v3.2.1, in which Markov Chain Monte Carlo samplings were run for 500,000 generations, with trees sampled every 1,000th generation. To define the phylogenetic groups by the eBURST analysis, we used data obtained from 750 samples including all 273 samples from Canada from this and previous studies [26, 30], 477 samples of *B. burgdorferi* collected from questing ticks and/or from ticks on hosts in the USA, and 19 samples identified to date only in human patients in the USA. All these data are freely available in the pubmlst.org database. From these samples, 138 unique STs were available for

use in the phylogenetic analyses. To support interpretation of the rooted phylogenetic tree in terms of the recent diversification of the different STs, a minimum spanning tree (MST) was constructed using goeBURST. The goeBURST analysis uses the distinct numerical allelic profile of each ST to reconstruct the phylogenetic links between genotypes and to infer an evolutionary descent pattern [47]. The eBURST algorithm within goeBURST defines the primary founder ST in the group of STs as the ST that has the greatest number of single-locus variants within the population of STs [44]. For this analysis the strength of the link between two STs is given as the SEB value (obtained up to triple locus variants: TLV), and the number of locus differences between them.

The intergenic spacer 16S-23S rrs-rrlA was amplified from the samples by nested PCR using the outer primers PA Forward (GTATGTTTAGTGAGGGGGGTG position: 2306–2326) and P95 Reverse (GGATCATAGCTCAGGTGGTTAG position: 3334–3313), and the inner primers PB Forward (AGGGGGGTGAAGTCGTAACAAG position: 2318–2339) and P97 Reverse (GTCTGATAAACCTGAGGTCGGA position: 3305–3284) as described in [42]. For comparison of IGS sequences from samples in this study with those from previous studies, we classified our samples according to three previously-used methods. First, IGS sequence type identification numbers were assigned according to the scheme of [42] by comparing the sequences from our samples to reference sequences available in Genbank (accession numbers AY275189 to AY275212). Second, we assigned expected Ribosomal Sequence Types (RSTs) as reported in [48] because this is frequently used to distinguish (in broad terms) genotypes of *B. burgdorferi* that have ecological differences and vary in pathogenicity [31, 48]. To do this we assigned ribosomal spacer identification numbers (RSPs) to deduced IGS types according to the method of [49] by comparing our sequences with the relevant reference sequences in Genbank (accession numbers: EF649781 for RSP1, EF649783 for RSP3, EF649784 forRSP4, EF649786 for RSP6, EF649787 for RSP7, EF649789 for RSP9, EF649790for RSP10, and EU477177 to EU477185 for RSP12 to RSP20). Then, again following [25], RST numbers were assigned to the following RSPs [25]: RST1 corresponding to RSP1 and RSP7; RST2 corresponding to RSP3, RSP4 and RSP20; and RST3 corresponding to RSP14, RSP9 and RSP18, RSP10, RSP12 and RSP13, RSP19 (Table A in S1 File). An unrooted Bayesian phylogenetic tree was constructed from the 80 available IGS sequences from the samples

investigated here, with 16 reference ribosomal sequence spacers (RSP) and 24 IGS type and subtype sequences downloaded from GenBank, using MrBayes.

The ospC gene was also amplified by semi-nested PCR using the outer primers OC6 (+) (AAAGAATACATTAAGTGCGATATT) and 623 (-) (TTAAGGTTTTTTTTGGACTTTCTGC), and the inner primers OC6 (+Fluo) (Fluorescein-AAAGAATACATTAAGTGCGATATT) and 602 (-) (GGGCTTGTAAGCTCTTTAACTG) as reported in Qiu et al. 2002 [43]. The ospC major groups were identified by multiple alignments performed in ClustalW2 using default settings [50]. Pairwise alignment was done using reference sequences downloaded from GenBank (accession numbers are EU482041 to EU482051 for ospC major groups A to K, EU375832 for ospC L, EU482052 and EU482053 for ospC M and N, EU482054 and EU482055 for ospC T and U, EF592542 for ospC B3, EU482056 for ospC E3, EF592547 for ospC F3, HM047876 for ospC X and HM047875 for ospC Y). The criteria for assigning ospC sequences to ospC major groups were those described by [43] i.e. difference similarity of ≥ 99% to be included in a major group, and a similarity of ≤ 90% to be excluded from a major group.

Data analysis

Diversity of rodents and B. burgdorferi genotypes among study sites

With unequal sampling effort in different sites and regions, it is difficult to compare the diversity of *B. burgdorferi* genotypes and host species (using richness as an index) among different locations. However, development of individual rarefaction curves [51, 52], is a commonly used method for comparing species richness among samples with unequal sampling effort [53]. This approach allows comparison between different communities/locations after each community/location is "rarefied" back to an equal number of sampled specimens [51, 54, 55]. Three analyses were performed using PAST version 2.17c [56]. These included analysis of host species richness by geographic region, and *B. burgdorferi* genotype richness by host species and region. For this analysis, *B. burgdorferi* genotypes were MLST sequence types (STs) [57]. With the exception of the analysis of host species richness, in these and subsequent analyses we considered that STs in questing ticks as well as STs obtained from hosts (directly from infected host tissues or from one engorged tick

collected from the host) as comprising different 'ecological sources' of *B. burgdorferi* genotypes. In doing so, we recognized that the frequency of STs in questing ticks is the product of the transmission of *B. burgdorferi* genotypes from all species of the tick-host community in a particular location (see below).

In order to compare the abundance and the frequency of *B. burgdorferi* genotypes in broad terms among the field sites and animal hosts, the observed distributions of these STs among hosts and regions were compared against their expected distributions (i.e. the mean value of the number of identified STs among the host species and/or geographic regions). This comparison was performed using the chi-squared goodness of fit test where the null hypothesis was that the abundance of genotypes is the same among different host species. The alternative hypothesis is that the STs are non-randomly distributed among hosts suggesting possible associations with host species. The test was conducted with 95% confidence intervals using Fisher's exact test by the Monte Carlo Estimation algorithm in SAS version V9.4 (SAS Institute Inc., Cary, NC, USA).

Borrelia burgdorferi genotype-host species associations

Samples used in this analysis were one positive engorged tick or tissue sample per infected host, as well as questing ticks collected in the locations where the small mammal trapping was conducted. The null hypothesis was that the proportion of samples positive for a particular genotype would be the same amongst sources (questing ticks and different species). Questing ticks were included in this analysis as a category because the frequency of different genotypes in each trapping location to which hosts are exposed equals the prevalence in the questing tick population, so it would be expected, in the absence of genotype-host species associations, that frequencies of genotypes in questing ticks and in hosts would be similar.

First, correspondence analysis was performed using SPSS V17 (SPSS Inc., Chicago, US) to explore the relationship between *B. burgdorferi* genotypes and host species/source and their geographic locations. The analysis was conducted for different levels of clonal complex inference (i.e. single locus variant and double locus variant).

The correspondence analysis informed the development of logistic regression models to individually assess associations between genotypes of *B. burgdorferi* (determined by clonal

complexes, ospC major groups, and RSTs) and host species/source. Mixed effects generalized linear models with a logit link function were developed in SAS 9.4, where the fixed effects were host species/source, while the number of sampling visits (as a category rather than as a continuous variable) and geographic region of origin with nested individual site ID numbers were both considered as random effects in the same models. Site ID nested by region was included as a random effect to account for regional and inter-site variations that were not explicitly explored, while the number of visits (as a category) was explored as a random effect as different numbers of visits per site may have been associated with different probabilities of finding genotypes by (for example) reflecting different seasons of sampling. For this purpose, the GLMM (Generalized linear mixed model) with GLIMMIX procedure was performed in SAS version 9.4.

The general models were structured as follows:

$$CCi = \beta_0 + \beta_1 + (visit|region/siteID)$$

Where CC_i is the clonal complex/*ospC* or RST type as a binary outcome (*i.e*. 0 = absence and 1 = presence); β_0 is the estimate of the occurrence of the CC_i when all covariates are equal to zero; β_1 represents the fixed effects for the host species/source and the random effects of the number of site visits and the site ID, nested by region, indicated in brackets.

The models were fitted using a non-blocked covariance matrix assumption and parameter estimates were obtained using Restricted Maximum Likelihood to avoid certain deficiencies of the Maximum Likelihood method which does not take into account the loss in degrees of freedom due to the use of the fixed effect estimator [58]. The robust standard error estimator (empirical 'sandwich' estimators in the GLIMMIX procedure of SAS) was used to ensure results using small sample sizes were more robust [59]. A backward elimination process was used to group host species that were not significantly different, although questing ticks remained the reference host/source throughout. To further explore the significance of findings,

minimal models were compared statistically against intercept models, and analyses were recreated in R version 3.2.3 (The R Foundation for Statistical Computing, Vienna, Austria) using logistic regression to see if statistical significance remained using different model constructions. When significant differences in genotype occurrence amongst hosts were found when including data from questing ticks in statistical models, the analyses were repeated without questing tick data to see if these associations remained significant to provide more robust evidence of host-genotype associations. The level of significance was $P < 0.05$.

Results

Genotyping of B. burgdorferi

Of the samples collected at the study sites, 437 were analyzed by MLST (for which 273 were successfully sequenced with 34 samples being rejected as having mixed infections), ospC (for which 240 were successfully sequenced with 25 samples being rejected as having mixed infections), IGS (for which 258 were successfully sequenced with 30 samples being rejected as having mixed infections). Following removal of multiple samples from hosts there were 91 samples (41 from hosts and 50 from questing ticks) for statistical analyses of host-genotype associations.

The eBURST and goeBURST analyses, using all 750 samples of the full MLST data set, identified 25 clonal complexes and 44 singletons using the SLV criterion, and 21 clonal complexes and 20 singletons using the DLV criterion. The 273 samples from Canada comprised 61 STs occurring in 18 different clonal complexes with 20 being singletons using SLV criterion, and 15 clonal complexes and 9 singletons using the DLV criterion. Using the SLV criterion, the largest clonal complex was CC34 which contained 9 STs (ST14, ST52, ST301, ST315, ST523, ST525, ST638, ST640, ST642), followed by CC12 which contained 4 STs (ST12, ST221, ST527, ST643), and CC4 and CC36 which both contained 3 STs (ST4, ST32, ST639, and ST9, ST36, ST537 respectively). The rest of the clonal complexes were minor complexes; each minor complex contained two sequence types (i.e. 12 STs and 10 others STs linked at least one ST from USA) (Fig 2).

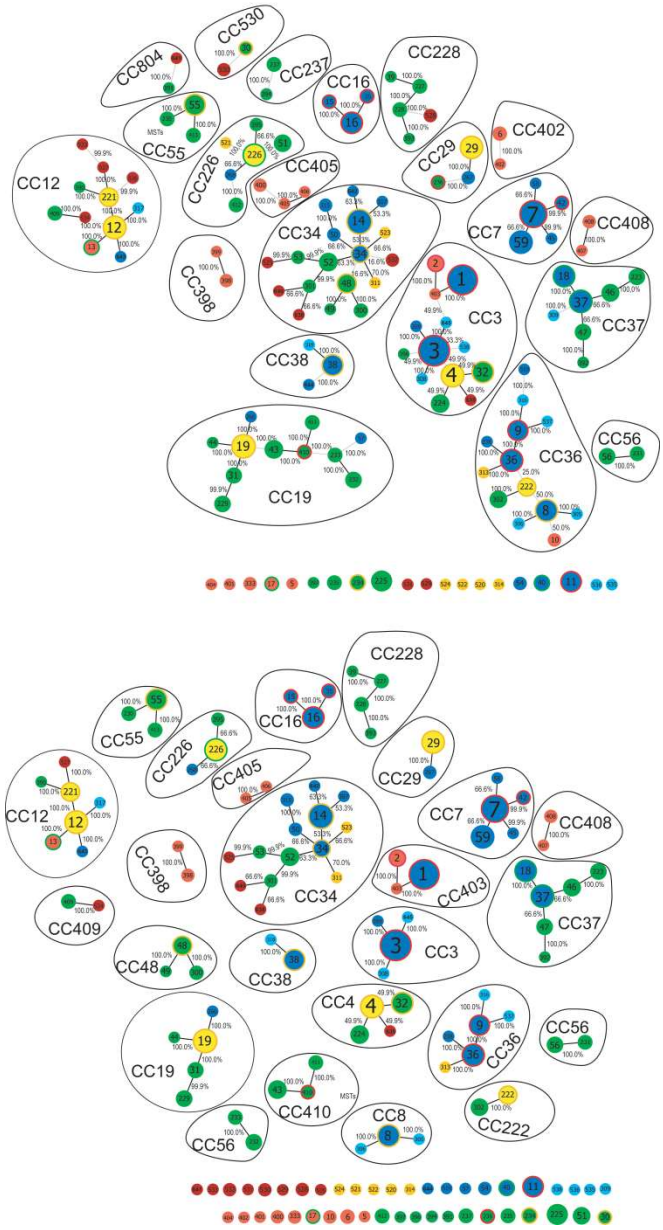


Figure 2: MLST clonal complexes.
Clonal complexes and singletons identified by eBURST and goeBURST analyses constructed at SLV (A) and DLV (B) levels. The spanning edge betweenness value (expressed as %) of the optimal edge is reported. STs are color-coded according to their geographic location: blue, STs found in the 'northeast' (*i.e.* STs found in Northeastern US and also in Quebec, eastern Ontario and the Maritimes); green, STs found in the 'Midwest' (*i.e.* STs found in Midwestern US and also in Manitoba); yellow, STs occurring in both the 'northeast' and the 'Midwest'; red, STs found in California; cyan, STs found only in the Maritimes; orange, STs found only at Long Point Ontario; brown, STs found only in Manitoba.

Using the DLV criterion, CC34 was again the largest clonal complex with 12 STs, followed by CC3, CC12, CC36 each with 7 STs, CC19 with 5 STs, and CC226 and CC228 each with 3 STs. The MSTs statistics (SEB) reported in the goeBURST diagram for edges between STs constituting each clonal complex using SLV (Fig 2A) and DLV (Fig 2B) criteria indicate that within clonal complexes the STs are highly related (i.e. the frequency with which they formed optimal edges in the MSTs forest tree for each CC was between 99.9% and 100%) and correspond mostly to significant (100% of the posterior probability) clades in the unrooted phylogenetic tree (e.g. CC19, CC55, CC38) (Fig A in S2 File). However, for certain large clonal complexes including CC34 some of the STs (particularly when using DLV criteria) had MST statistics indicating lower relatedness with other clonal complex members (with frequencies of the optimal edges ranging from 16% to 66.6%) and these STs frequently came from different clades in the phylogenetic tree. An example is CC36, which using the DLV criterion, comprises 3 groups of STs from three distinct (with 94% the posterior probability) clades of the unrooted phylogenetic tree (Fig A in S2 File) that are linked by edges reproduced only in ≤ 50% cases in the MST forest tree (Fig 2).

Intergenic spacer sequences were obtained from 80 of the samples used in statistical analyses and comprised 7 IGS types and 5 IGS subtypes (Table B in S1 File). An unrooted phylogenetic tree of these sequences (Fig 3) showed that the sequences and 12 IGS types and subtypes are well clustered into the three distinct ribosomal sequence types, RST1, RST2 and RST3. RST1 contains two IGS types (1 and 3) and one subtype (1A), RST2 contains one IGS type (4) and three subtypes (2A, 2D and 4A), and RST3 was the largest group comprising two IGS types (5 and 9) and four subtypes (6A, 6B, 7A and 8C).

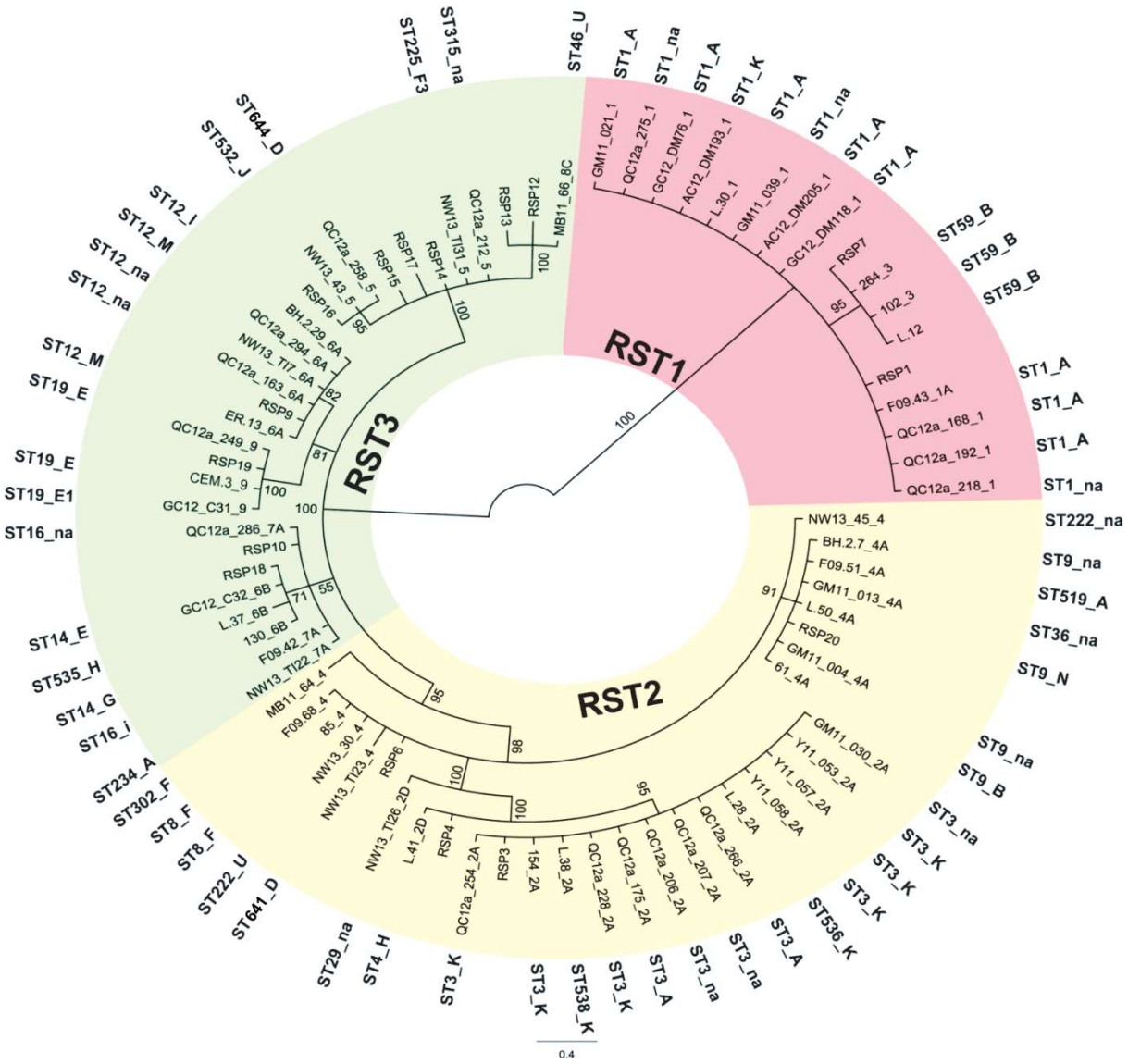


Figure 3: An unrooted Bayesian maximum likelihood tree of ribosomal 16S-23S rrs-rrlA IGS spacer sequences.

The tree was constructed using the 80 samples for which these sequences were available, as well as reference sequences described in the text. Identification numbers of the samples and the IGS type number are shown on the tree. The correspondence of these sequences to RST types is shown by colour coding (RST1: dark pink; RST2: light pink; RST3: gray). Around the colour-coded areas are the corresponding ST numbers and ospC major groups of the samples (na = data not available).

Alleles belonging to 15 major ospC groups (including A, B, D, E, E1, F, F3, G, H, I, J, K, M, N and U) were identified among 69 samples used in statistical analyses for which full ospC sequence data were available. The most frequent major group found was ospC A, which corresponded to 4 STs (ST1, ST3, ST9, ST519), ospC K corresponded to ST1, ST3, ST536 and ST538, and other ospC major groups were linked to one or two STs (Table B in S1 File). The ospC major groups were associated with different RSTs with 80% of ospC A being associated with RST1 and 20% associated with RST2, 75% of ospC B being associated with RST1 and 25% associated with RST2, while 94% of ospC K were associated with RST2 and 6% were associated with RST1. ospC major groups E, E1, G, J and M were all associated with RST3 (Table B in S1 File, and Fig 3).

Diversity of rodents and B. burgdorferi genotypes among study sites

Details of the origin of the samples and the *B. burgdorferi* ST frequencies are shown in Table 1. The rarefaction curves suggested that mammal species richness was similar among the regions, and that detected specific richness would rise at approximately even rates with increased sampling effort, and plateau at approximately the same sample size, in each region (Fig 4).

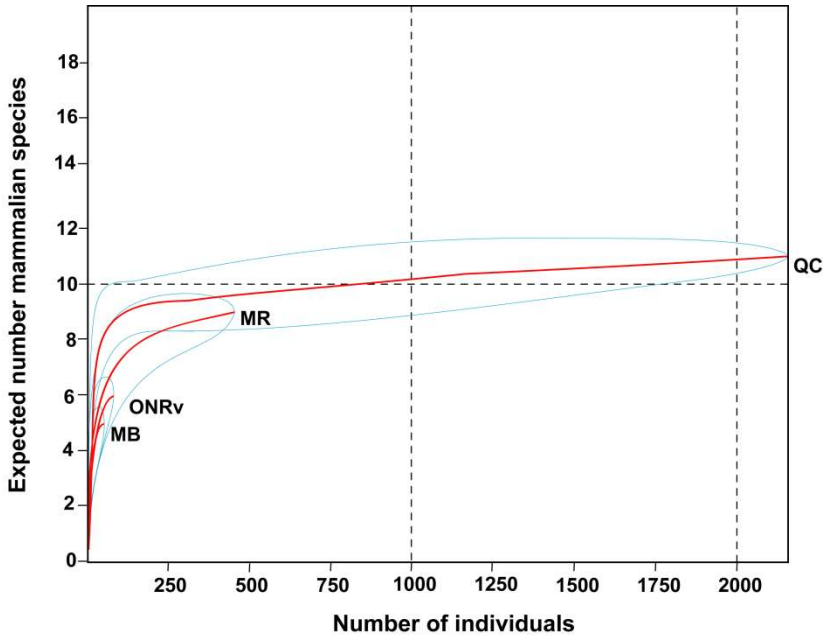


Figure 4: Rarefaction curves of rodent species richness.
Comparisons are made among the four geographic zones where rodent trapping was conducted (Manitoba [MB], Ontario Rainy River [ONRv], Quebec [QC] and the Maritimes [MR]) in the richness of rodent communities using rarefaction measurements. Blue lines depict the 95% confidence limits.

In contrast, among the regions the highest ST richness was in Manitoba (33 STs) followed by Quebec (15 STs) and the Maritimes (15 STs), and the individual rarefaction curves suggested that the detected richness of STs would rise faster by increased sampling effort in Manitoba compared to other regions (Fig 5a).

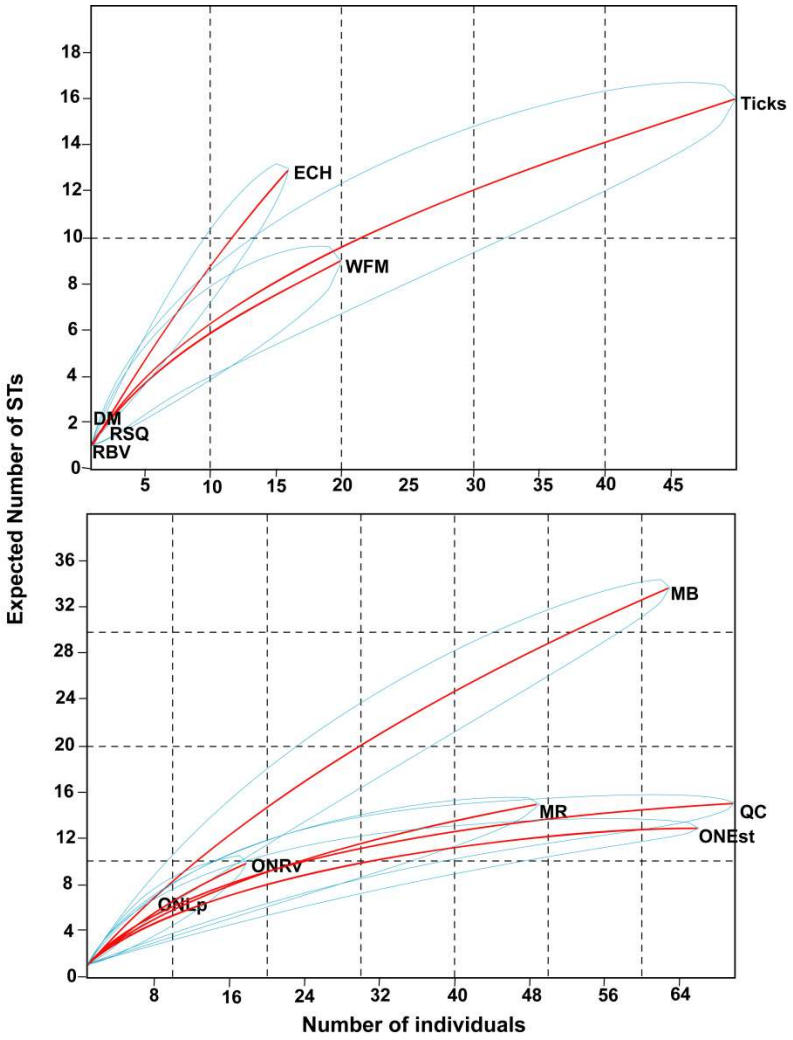


Figure 5: Rarefaction curves of ST richness.
Comparisons of the richness of *B. burgdorferi* STs using individual rarefaction curves among four different geographic regions (A) and the six ecological sources (B). The geographic regions are Manitoba [MB], Ontario at Rainy River [ONRv], Quebec [QC] and the Maritimes [MR]. The sources are Deer mouse (DM), Eastern chipmunk (ECH), Red-backed vole (RBV), Red squirrel (RSQ), White-footed mouse (WFM), and questing ticks (Ticks). Blue lines depict the 95% confidence limits.

The richness of STs among host species was highest in the eastern chipmunk with 24% of the total STs (13 STs) followed by white-footed mice with 16% of STs (9 STs), which had similar specific richness of STs to host-seeking ticks. The individual rarefaction curves suggested that the richness of STs would rise faster by increasing sampling effort in eastern chipmunks compared to increased sampling of other rodents or questing ticks (Fig 5b). The chi-square goodness of fit test suggested that *B. burgdorferi* genotypes were non-randomly distributed

among hosts/sources ($\chi 2 = 935.38$, df = 6, $P < 0.001$) and among geographic locations ($\chi 2 = 78.63$, df = 5, $P < 0.001$) (Table 2). Note that the level of statistical significance was adjusted according to Bonferroni correction to $P < 0.002$ (= 0.05/21) for the fixed factor host species/ticks and $P < 0.003$ (= 0.05/15) for the fixed factor geographic location, to account for multiple one-way comparisons. Thus, this analysis suggests that there were significant associations between *Borrelia* genotypes and host species/ticks and between *Borrelia* genotypes and geographic locations. The Mantel-Haenszel chi square test suggests that the relationship between STs and geographic zones was linear ($\chi 2 = 8.93$, df = 1, $P = 0.002$) supporting the specificity of certain genotype ranges to some locations (Table 2). In both cases the contingency coefficient (Cf) suggests that the genotypes are strongly associated with certain hosts/sources (Cf = 0.80) and geographic locations (Cf = 0.81) (Table 2).

Table 2: Chi-square goodness of fit test results. Results for associations among hosts and clonal complexes (CCs), and among locations and CCs are shown.

Host-CCs			
Statistic	DF	Test value	P
Chi-Square	6	935.38	< 0.001
Monte Carlo Estimate for the F Exact Test			0.001
Mantel-Haenszel Chi-Square	1	2.69	0.101
Monte Carlo Estimate for the F Exact Test			0.102
Contingency Coefficient		0.80	
Locations-CCs			
Statistic	DF	Test value	P
Chi-Square	5	78.63	< 0.001
Monte Carlo Estimate for the F Exact Test			< 0.001
Mantel-Haenszel Chi-Square	1	8.93	0.002
Monte Carlo Estimate for the F Exact Test			0.002

Borrelia burgdorferi genotype-host species associations

For this analysis there were 91 samples comprising 30 STs that, in the goeBURST analysis described above, fell into 15 of the clonal complexes with 8 that were singletons using the SLV criterion (Table C in S1 File), or fell into 11 of the clonal complexes with 5 singletons using the DLV criterion (Table D in S1 File). The correspondence analysis shows evidence of

an association between certain host species and clonal complexes suggesting that significant host-genotype associations exist, regardless of whether clonal complexes were formed using SLV criteria ($\chi 2 = 158.28$, df = 110 and P = 0.002) or DLV criteria ($\chi 2 = 108.14$, df = 75 and P = 0.007) (details in Tables E to J in S1 File). The two-dimension solution of the principal component analysis explains 66.7% of the variation when clonal complexes are developed using the SLV criterion (Fig 6A) and 83.2% when clonal complexes were developed using the DLV criterion (Fig 6B) indicating that most of the variation reflected an association between host species and clonal complexes. In both cases, correlation of CC34 and chipmunks is high (Table F and J in S1 File). Correspondence analysis also suggested associations between the white-footed mouse and CC403. The analysis suggested an association between red squirrels and clonal complexes although there were only two individuals of this species in the data set (Table 1, Table F and J in S1 File, Fig 6).

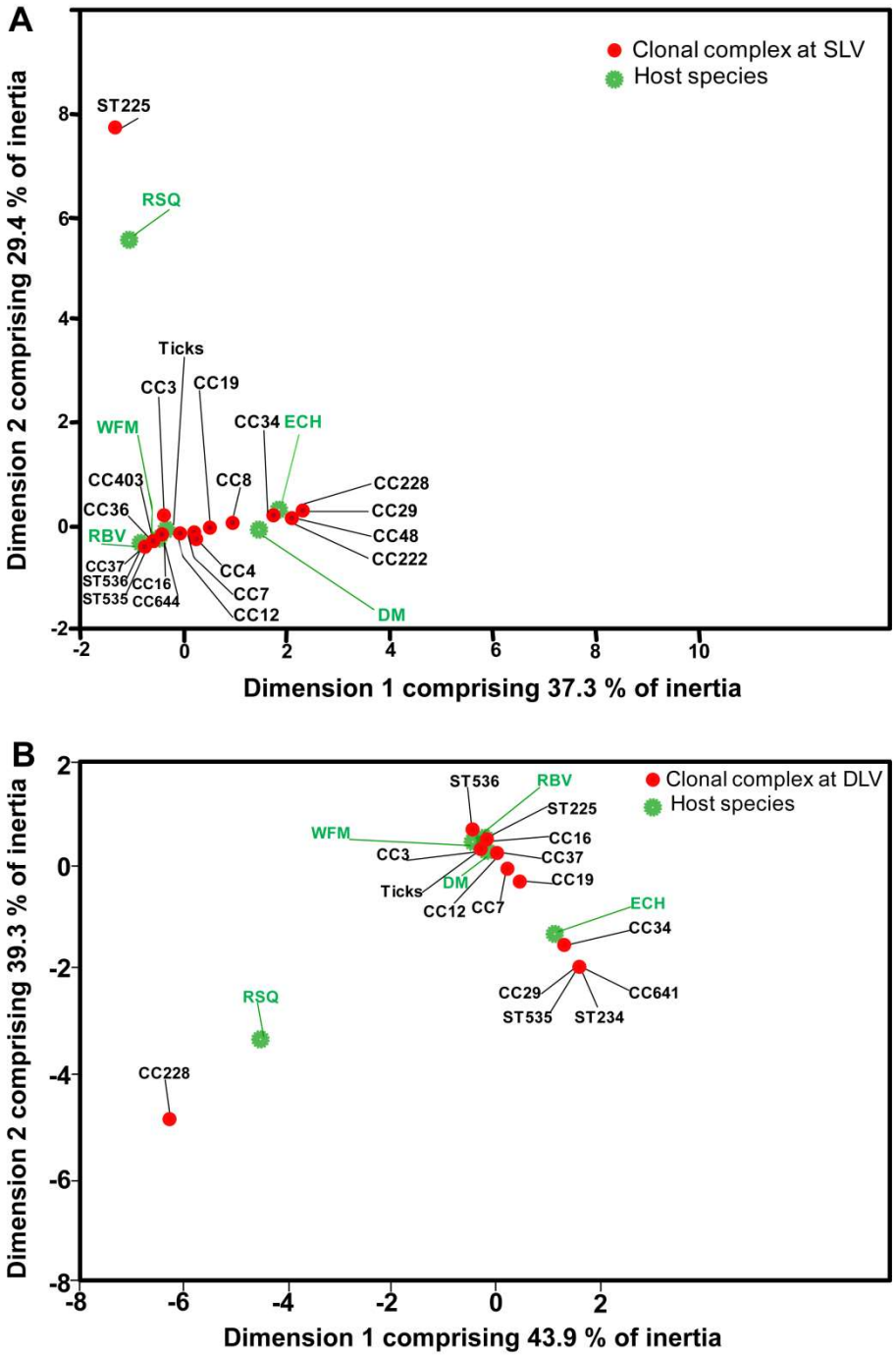


Figure 6: Correspondence analysis results.
Correspondence analysis biplot maps are shown for associations among host species/sources and the MLST clonal complexes when these were obtained using SLV (A) and DLV (B) criteria.

Results of the GLIMMIX model analyses are summarised in Table 3 and detailed in Tables K to T in S1 File. Significant associations were found between chipmunks and STs of CC34 (when constructed by both SLV and DLV criteria) and with RST2 type IGS sequences (IGS4) and ospC G. Significant associations were found between white-footed mice and CC403 (which was only present when clonal complexes were constructed with SLV criteria), RST1 type IGS sequences and ospC A. Significant associations were also found between deer mice and CC4 (which was only present when clonal complexes were constructed with SLV criteria), and with ospC H. In each case the prevalence of these STs and IGS and ospC sequences were significantly different from the prevalence in questing ticks and in other host species (Tables K to R in S1 File). Also in each case the significance of the minimal models was supported by being significantly different from the intercept-only model (Table T in S1 File). When the minimal models were reconstructed in R software, all remained significant (with and without questing tick data: Tables U and V in S1 File) except for the associations between white-footed mice and CC403 and between white-footed mice and RST1 type sequences, which were marginally non-significant (P = 0.097 and 0.056 respectively) in models with questing tick data included.

Table 3: Significant associations of host species with different genotypes of *B. burgdorferi*. Host species abbreviations are: DM = deer mouse, ECH = eastern chipmunk and WFM = white-footed mouse. * indicates that the random effect of site ID nested by region was significant and included in the model while ** indicates that the random effect of the number of site visits was significant and included in the model.

Factor	Estimate	Standard Error	DF	t Value	P > |t|
		CC34S*			
Intercept	0.01298	0.07381	22	0.18	0.862
ECH	0.3207	0.09097	66	3.53	0.001
Other host spp.	-0.0146	0.07325	66	-0.20	0.843
		CC34D*			
Intercept	0.01951	0.07786	22	0.25	0.804
ECH	0.4083	0.09428	66	4.33	<0.001
Other host spp.	-0.0287	0.07564	66	-0.38	0.705
		CC403			
Intercept	0.1800	0.05411	88	3.33	0.001
WFM	0.3500	0.08555	88	4.09	<0.001
Other host spp.	0.04762	0.08349	88	0.57	0.570
		CC4			

Intercept	0.06157	0.06311	22	0.98	0.340
DM	0.4662	0.1296	66	3.60	0.001
Other host spp.	-0.0279	0.05841	66	-0.48	0.633
RST 1 type IGS sequences*					
Intercept	0.1905	0.06084	77	3.13	0.002
DM	0.3500	0.08817	77	3.97	<0.001
Other host spp.	0.05556	0.09293	77	0.60	0.552
RST 2 type IGS sequences (IGS4)					
Intercept	0.02381	0.03814	77	0.62	0.534
ECH	0.2839	0.07846	77	3.62	<0.001
Other host spp.	0.01619	0.06244	77	0.26	0.796
RST 2 type IGS sequences (IGS2D)					
Intercept	0.02381	0.03226	77	0.74	0.463
DM	0.4762	0.1513	77	3.15	0.002
Other host spp.	0.03175	0.04748	77	0.67	0.506
ospC G**					
Intercept	-0.00216	0.04874	32	-0.04	0.965
ECH	0.2471	0.07165	34	3.45	0.001
Other host spp.	-0.00637	0.05550	34	-0.11	0.909
ospC A					
Intercept	0.2632	0.06698	66	3.93	<0.001
WFM	0.3529	0.1001	66	3.52	0.001
Other host spp.	-0.2632	0.1291	66	-2.04	0.050
ospC H					
Intercept	0.02632	0.03647	66	0.72	0.473
DM	0.4737	0.1631	66	2.90	0.005
Other host spp.	0.04265	0.05543	66	0.77	0.444

Phylogenetic analysis of the host-genotype associations

The rooted Bayesian phylogenetic tree with outgroups (Figs 7 and 8) shows that many clonal complexes (e.g. CC37, CC403, CC226, CC16) are correlated with clades. However, certain clonal complexes such as CC34 do not form clear clades. For CC34 some STs (ST52, ST53, ST301, ST525, ST638, ST640) form a clear clade with 100% posterior probability, while the rest (e.g. ST14, ST48, ST300, ST532) group at the base of the tree (Fig 7). For clarity, the relationships of STs among CCs 34, 4 and 403 in a phylogenetic tree constructed using only the STs that are members of these CCs are shown in Fig 8.

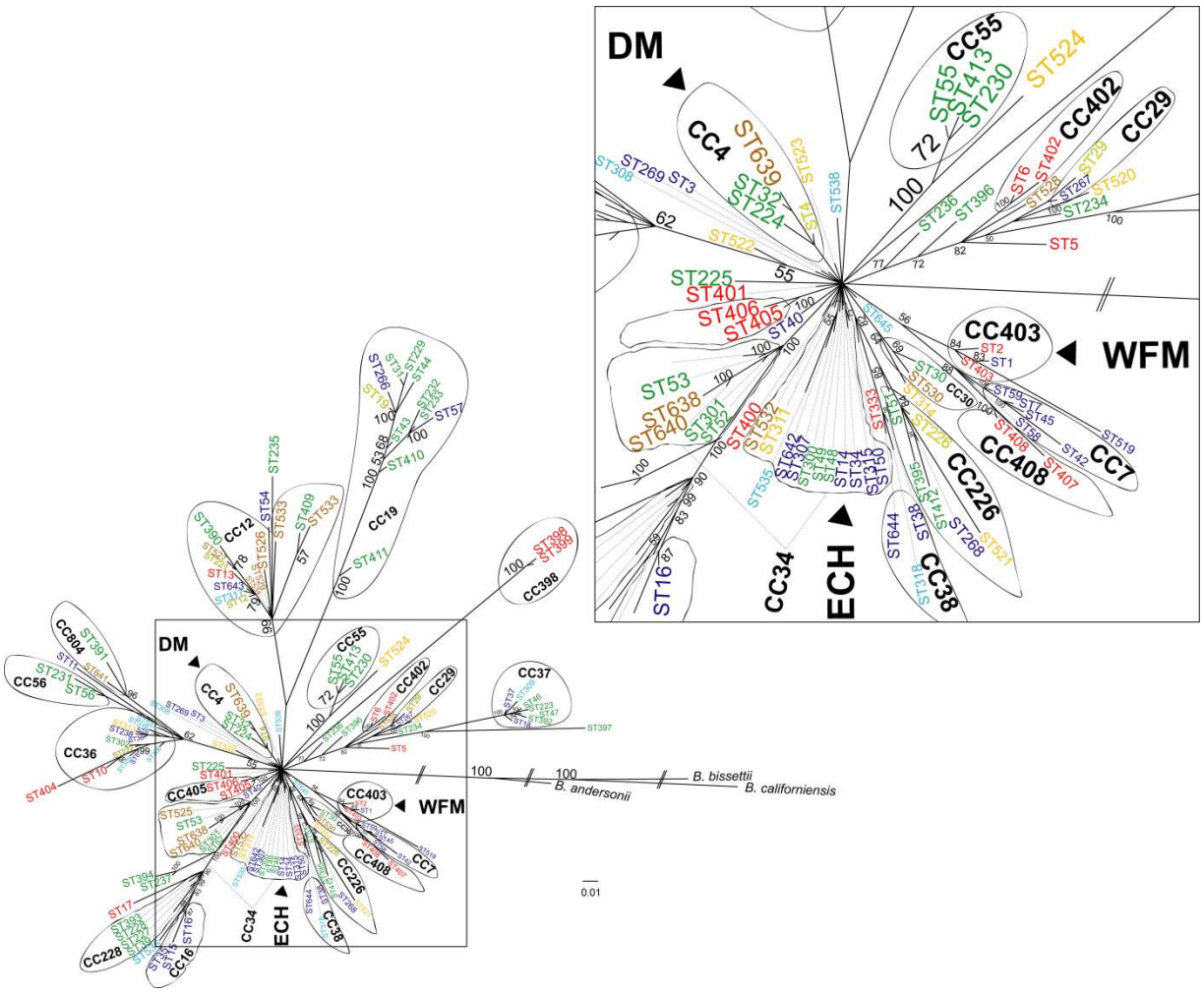


Figure 7: A rooted Bayesian phylogenetic tree of MLST STs.
STs are color-coded according to their geographic location: cyan: Maritime Provinces, orange: Long Point Ontario, brown: Manitoba, blue: Northeastern USA, eastern Ontario and southwestern Quebec, green: Midwest USA, yellow: STs found in both northeastern and Midwestern USA, and red: California. Posterior probabilities are shown beside nodes. The scale bar corresponds to the number of substitutions per unit branch length. STs of clades that belong to distinct clonal complexes are encircled and numbered as in Fig 2. ECH, WFM and DM indicate STs of clonal complexes associated with chipmunks, white-footed mice and deer mice respectively. Outgroups are *B. bissettii*, *B. andersonii* and *B. californiensis*. An enlargement of the central part of the tree is shown to the right.

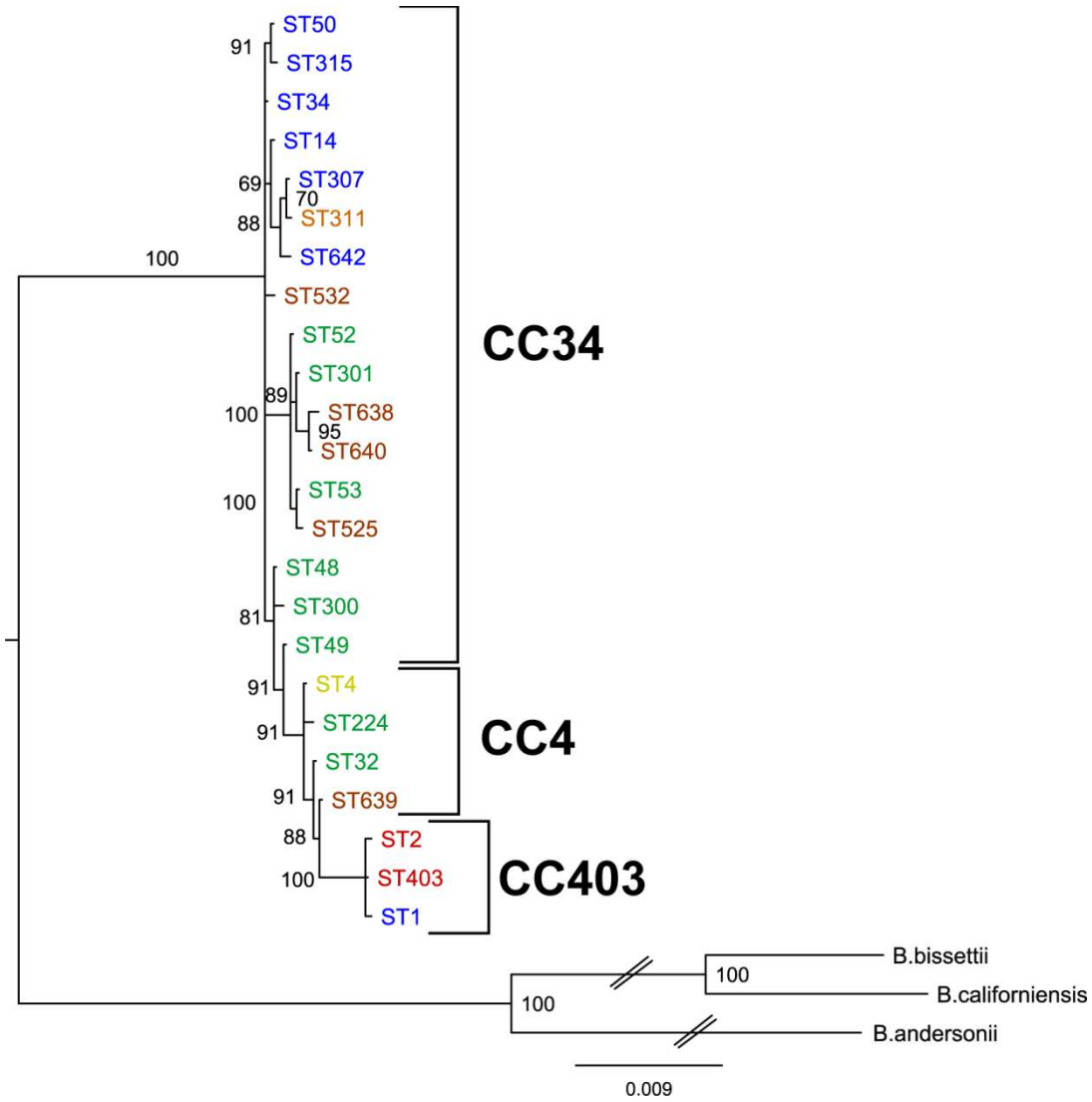


Figure 8: The phylogenetic relationship of STs of clonal complexes associated with rodents. This shows a phylogenetic tree constructed using the same method and the same outgroups as the tree in Fig 7, but with only the STs of the rodent-associated clonal complexes CC34, CC4 and CC403.

The association of eastern chipmunks with CC34 was due to associations with ST14, ST300 and ST532, which are linked in one part of CC34 (Fig 2). There was evidence of associations between deer mice and white-footed mice respectively with CC4 and CC403, STs of which form (on the basis of branch length) more recently evolved clades of the phylogenetic tree (Fig 8). The MST developed using goeBURST shows the optimal parent-descendent linkage among the STs (Fig B in S2 File). For the SLV, DLV, and TLV criteria, ST34 is always predicted as the founder ST with a maximum bootstrap value of 83% obtained at SLV. This

means that while the SEB statistics between ST34 (the overall founder and founder of CC34) and ST3 (the founder of CC3), and between ST3 and ST4 (the founder of CC4) and ST1 are low (possibly suggesting horizontal gene transfer: [445]), ST34 may be ancestral to ST3 of CC3, which in turn is most likely to be ancestral to ST1 of CC403 and ST4 of CC4.

Discussion

To our knowledge, *B. burgdorferi* remains a generalist pathogen which can survive in and be transmitted from many host species [4]. In this study we explored possible statistical associations between genotypes of *B. burgdorferi* with different host species. The samples available to us came from multiple studies and we used a number of techniques to carefully control for sampling effects. First, we used rarefaction curves to explore whether sites likely differed in the ranges of host species and *B. burgdorferi* genotypes, and whether the range of *B. burgdorferi* genotypes differed among host species/sources of *B. burgdorferi* DNA. This analysis suggested that sites sampled in the different geographic regions had similar host species richness, but *B. burgdorferi* genotype richness was higher in the Manitoba sites compared to sites in other regions. This result is consistent with previous studies in the USA that suggest that *B. burgdorferi* genotype richness is higher in the upper Midwest (i.e. immediately south of Manitoba) than in the northeast (i.e. immediately south of eastern Ontario, Quebec and the Maritimes) [14]. Greater seasonal synchrony of activity of larval and nymphal *I. scapularis* ticks may explain these geographic differences in *B. burgdorferi* genotype richness. Greater seasonal asynchrony of immature tick activity in the US northeast, with nymphs infecting hosts when they are active in spring to early summer and larvae acquiring infections possibly some months later in late summer, has been associated with higher frequencies of genotypes that have putatively long-lived infections in reservoir hosts [60]. This is consistent with the idea that seasonal asynchrony selects for such genotypes [4, 61]. It has been hypothesised that greater persistence of transmissible infections in the host requires greater adaptation to particular host species [4, 61], so seasonal synchrony of immature ticks in the US Midwest may permit transmission of a wider diversity of less host-adapted genotypes than in the northeast. The rarefaction curves suggest some differences in

richness of genotypes among host species, which was supported by the abundance analysis conducted by the goodness of fit statistic. In particular genotype richness was highest in chipmunks and this could be due to the relatively long life (up to three years: [62]) of this species compared to the other dominant rodent species such as white-footed mice which rarely survive for a year [63]. The longer a host survives, the more *B. burgdorferi*-infected tick bites it will receive so, with the often persistent nature of *B. burgdorferi* infections, longer lived species would be expected to exhibit a wider range of genotypes on the assumption that genotypes are not absolute host specialists. Furthermore, chipmunks may carry higher numbers of nymphal ticks than mice, thus further enhancing their capacity to be infected with a higher diversity of genotypes [14].

The possibility of associations among host species and *B. burgdorferi* genotypes was initially explored by correspondence analysis, which suggested that there were non-random associations of genotypes among host species. This was then further explored by generalised linear models which demonstrated, in separate analyses, the following associations: i) one clonal complex (CC34 using both SLV and DLV criteria), RST type 2 IGS sequences (particularly IGS type 4) and ospC major group G with chipmunks; ii) one clonal complex (CC4) and ospC major group H with deer mice and iii) one clonal complex (CC403), RST type 1 sequences and ospC major group A with white-footed mice. The analyses accounted for geographic location, uneven sampling among sites, and small sample sizes. The associations of *Borrelia* genotypes with chipmunks were robust throughout the study, and chipmunk samples were available from nearly all geographic regions (Manitoba, northwestern Ontario, southeastern Ontario-southern Quebec and the Maritimes). The mouse-genotype associations were based on limited sample sizes, and samples from deer mice and white-footed mice were limited in their geographic distribution (deer mouse samples from Manitoba and northwestern Ontario, and white-footed mouse samples from Quebec and the Maritimes). Two of the associations with white-footed mice (CC403 and RST1 sequences) were marginally non-significant when models were run in R when questing tick data were included. Furthermore, although the great majority of our carried single (or single dominant) genotype infections, information from mixed genotype infections was excluded. It would be expected that if hosts are bitten by ticks with mixed-genotype infections then, unless there is some form of selection

in the host (other than the "first arriving genotype taking all" principle [64]) frequencies of genotypes in mixed-genotype infections and single-genotype infections should be similar although more research on mixed-genotype infections is needed. Therefore, further explorations of these host-genotype relationships are warranted. Nevertheless, some associations (e.g. associations of CC403 with white-footed mice) were consistent for logistic regression and correspondence analysis, and associations of ospC A sequences with white-footed mice and ospC G with chipmunks are consistent with other studies [24]. It was not surprising that all three genotyping methods (MLST, ospC and RST types) found associations between host species and *Borrelia* genotypes. In previous studies of the genetic diversity of *B. burgdorferi* using the same typing methods, it has been found that ospC major groups are often associated with the same STs [41] and there is an overall association between ospC major groups and rrs-rrlA IGS RSTs (e.g. ospC A and B with RST1: [65]), due to relatively high levels of linkage disequilibrium. Here, there was partial evidence of linkage disequilibrium with the STs, IGS types and ospC major groups associated with the same host species being found in the same samples in many cases, but not in all (see Fig 4). Samples with ST1 frequently carried ospC major group A and IGS sequences were of type 1, all of which were associated with white-footed mice, but some samples with ST1 and IGS type 1 had ospC sequences of major group K. Samples with ST14 (of CC34) also carried ospC major group G, but not IGS group 4, when all were associated with chipmunks. The one sample from deer mice with all three sequences was ST4, IGS type 2D and carried ospC major group H.

There was no evidence in our data for the type of near-complete host specialisation of clonal complexes that is seen with European *Borrelia* genospecies, which involves a well-documented mechanism of genospecies-specific sensitivity to the alternative pathway of complement of different host species [15]. Clear one-to-one host species ospC group associations would not be expected as this is not seen in European genospecies [66]. While the prevalence of STs of CC34 (using DLV criteria) in chipmunk samples was 31% (5/16) compared to 2% (1/50) in questing ticks, 69% of genotypes in chipmunk samples were of different clonal complexes. Similarly, the prevalence of STs of CC403 was 35% (7/20) in white-footed mouse samples and 18% (9/50) in questing ticks so 65% of STs in white-footed mouse samples were of different clonal complexes. Associations of genotypes of North

American *B. burgdorferi* are most likely due to characteristics of the genotypes of a more subtle nature that enhance the likelihood of finding them in, or being transmitted from, a particular host species. These may include particular susceptibility of the host species to the genotype, less pathogenic effects of the genotype in the host species (although pathogenicity in wild hosts is not a common feature [67]), longer persistence of infection and infectivity (i.e. transmissibility to feeding ticks) of the genotype in the host species [25], more efficient transmission of the genotype from the host species to feeding ticks, and possibly support of co-feeding transmission of ticks in addition to transmission from systemic host infections as previously reviewed [4, 61]. Further research is needed to elucidate the mechanisms that underlie these observed associations.

Hanincova et al [14] made comparisons as to the power of MLST typing versus ospC alleles to predict pathogenicity of *B. burgdorferi* genotypes, but here we do not make any comparisons, and simply identify that host-genotype associations were detected using the range of different methods for typing genotypes that have been applied to study the genetic diversity of *B. burgdorferi* [24, 41, 42, 43, 65, 68]. However, the concatenated housekeeping gene sequences used for MLST typing, which show stabilising selection and neutral variation, do provide more robust data for creating phylogenetic trees than ospC (which is under balancing selection) and perhaps IGS (which is not thought to be subject to selection: [1]). The rooted MLST phylogenetic tree suggested that STs associated with mice occurred in clades that (on the basis of branch length supported by the minimum spanning tree: Fig B in S2 File) may have evolved more recently than STs associated with chipmunks (Figs 7 and 8). We hypothesise that chipmunk-associated stains are more closely related to ancestral genotypes from which currently circulating chipmunk and mouse associated genotypes have evolved. This hypothesis may also be supported by several pieces of information on the role of chipmunks in transmission of *B. burgdorferi* and their current and past geographic distribution.

First, chipmunks can have an important role in the enzootic transmission cycle of *B. burgdorferi*; being important hosts for immature ticks, efficiently transmitting the bacterium to ticks, and often being considered as being second to mice in importance as reservoir hosts only

because of their lower relative density [69–75]. Therefore, it is possible that chipmunks could maintain transmission cycles of *B. burgdorferi* in the absence of mice, particularly for genotypes that may be more adapted to, and transmissible from them. Second, the hypothesis of adaptation to contemporary host species may be supported by the phylogeography of these small mammals. Several comparative studies have suggested that eastern chipmunk, white-footed mice and deer mice had a similar evolutionary history in northeastern and Midwestern North America since the retreat of the ice after the last glacial maximum approximately 20,000 years ago [76, 77]. However, evidence from fossil data [76, 78], suggests that eastern chipmunks were among the rare small mammal species to have survived and persisted in multiple refugia in northern locations during the last glacial period, and experienced a southward expansion towards the central USA when the ice retreated. Therefore, it is possible that populations of genotypes of *B. burgdorferi* were maintained by eastern chipmunks (with potentially host specific nidicolous ticks) in these refugia, and expanded when the climate became more favourable in the late Pleistocene for expansions and co-occurrence of tick vectors, *B. burgdorferi* populations, as well as other host species such as *Peromyscus spp*. mice [79–81]. Recent studies support the climate-sensitive nature of *P. leucopus* distributions [82].

In this study we have provided evidence of host association of genotypes of *B. burgdorferi*, and that these associations occur with genotypes that cluster phylogenetically. These findings support the view that the MLST-defined tree topology reflects both demographic processes (population expansions and contractions) and also associations of host species with *Borrelia* genotypes. Host adaptation may have been the driver for the differences in pathogenicity of North American *B. burgdorferi* genotypes that are also reflected by the phylogeny and evolutionary history of this bacterium [14]. Further research is needed to better describe the extent and strength of host-genotype associations, to reveal how these occur mechanistically, and to predict their consequences for human health.

Acknowledgments

This study was funded by the Public Health Agency of Canada.

Funding Statement

The study was funded by the Public Health Agency of Canada.

Data Availability

All relevant data are within the paper and its Supporting Information files.

References

1. Margos G, Vollmer SA, Ogden NH, Fish D. Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato. Infect Genet Evol. 2011; 11: 1545–1563. doi: 10.1016/j.meegid.2011.07.022 PMID: 21843658.
2. Gern L, Humair PF. Ecology of *Borrelia burgdorferi* sensu lato in Europe. In: Gray JS, Kahl O, Lane RS, Stanek G. Lyme *Borreliosis*: Biology, Epidemiology and Control. Wallingford: CABI Publishing; 2002. pp. 149–174.
3. Korenberg EI, Gorelova NB, Kovalevskii YV. Ecology of *Borrelia burgdorferi* sensu lato in Russia. In: Gray JS, Kahl O, Lane RS, Stanek G. Lyme *Borreliosis*: Biology, Epidemiology and Control. Wallingford: CABI Publishing; 2002. pp. 175–200.
4. Kurtenbach K, Hanincová K, Tsao JI, Margos G, Fish D, Ogden NH. Fundamental processes in the evolutionary ecology of Lyme *borreliosis*. Nat Rev Microbiol. 2006; 4: 660–669. PMID: 16894341 Strle F, Stanek G. Clinical manifestations and diagnosis of Lyme *borreliosis*. Curr Probl Dermatol. 2009; 37: 51–110. doi: 10.1159/000213070 PMID: 19367097.
5. Halperin JJ. Chronic Lyme disease: misconceptions and challenges for patient management. Infect Drug Resist. 2015; 8: 119–128. doi: 10.2147/IDR.S66739 PMID: 26028977.

6. Jungnick S, Margos G, Rieger M, Dzaferovic E, Bent SJ, Overzier E, et al. *Borrelia burgdorferi* sensu stricto and *Borrelia afzelii*: Population structure and differential pathogenicity. Int J Med Microbiol. 2015; 305: 673–681. doi: 10.1016/j.ijmm.2015.08.017 PMID: 26341331.
7. Piesman J, Gern L. Lyme *borreliosis* in Europe and North America. Parasitology. 2004; 129: 191–220.
8. Wormser GP, Dattwyler RJ, Shapiro ED, Halperin JJ, Steere AC, Klempner MS, et al. The clinical assessment, treatment, and prevention of Lyme disease, human granulocytic *anaplasmosis*, and *babesiosis*: clinical practice guidelines by the Infectious Diseases Society of America. Clin Infect Dis. 2006; 43: 1089–1134. PMID: 17029130.
9. Seinost G, Dykhuizen DE, Dattwyler RJ, Golde WT, Dunn JJ, Wang IN, et al. Four clones of *Borrelia burgdorferi* sensu stricto cause invasive infection in humans. Infect Immun. 1999; 67: 3518–3524. PMID: 10377134.
10. Dykhuizen DE, Brisson D, Sandigursky S, Wormser GP, Nowakowski J, Nadelman RB, et al. Short report: The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. Am J Trop Med Hyg. 2008; 78: 806–810.
11. Wormser GP, Brisson D, Liveris D, Hanincova K, Sandigursky S, Nowakowski J, et al. *Borrelia burgdorferi* genotype predicts the capacity for hematogenous dissemination during early Lyme disease. J Infect Dis. 2008; 198: 1358–1364. doi: 10.1086/592279 PMID: 18781866.
12. Wormser GP, Liveris D, Nowakowski J, Nadelman RB, Cavaliere LF, McKenna D, et al. Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. J Infect Dis. 1999; 180: 720–725. PMID: 10438360.
13. Hanincova K, Mukherjee P, Ogden NH, Margos G, Wormser GP, Reed KD, et al. Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. PLoS One. 2013; 8: e73066. doi: 10.1371/journal.pone.0073066 PMID: 24069170.
14. Kurtenbach K, De Michelis S, Etti S, Schäfer SM, Sewell HS, Brade V, et al. Host association of *Borrelia burgdorferi* sensu lato: the key role of host complement. Trends Microbiol. 2002; 10: 74–79. PMID: 11827808.

15. Margos G, Wilske B, Sing A, Hizo-Teufel C, Cao WC, Chu C, et al. *Borrelia bavariensis sp*. nov. is widely distributed in Europe and Asia. Int J Syst Evol Microbiol. 2013; 63: 4284–4288. doi: 10.1099/ijs.0.052001-0 PMID: 23838444.
16. Dsouli N, Younsi-Kabachii H, Postic D, Nouira S, Gern L, Bouattour A. Reservoir role of lizard *Psammodromus algirus* in transmission cycle of *Borrelia burgdorferi* sensu lato (Spirochaetaceae) in Tunisia. J Med Entomol. 2006; 43: 737–742. PMID: 16892633.
17. Hanincova K, Kurtenbach K, Diuk-Wasser D, Brei B, Fish D. Epidemic spread of Lyme *borreliosis*, northeastern United States. Emerg Infect Dis. 2006; 12: 604–611. PMID: 16704808.
18. Wood CL, Lafferty KD. Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. Trends Ecol Evol. 2013; 28: 239–247. doi: 10.1016/j.tree.2012.10.011 PMID: 23182683.
19. Sasal P, Niquil N, Bartoli P. Community structure of digenean parasites of sparid and labrid fishes of the Mediterranean sea: a new approach. Parasitology. 1999; 119: 635–648. PMID: 10633926.
20. Giorgi MS, Arlettaz R, Guillaume F, Nusslé S, Ossola C, Vogel P, et al. Causal mechanisms underlying host specificity in bat ectoparasites. Oecologia. 2004; 138: 648–654. PMID: 14714175.
21. Bush SE, Clayton DH. The role of body size in host specificity: reciprocal transfer experiments with feather lice. Evolution. 2006; 60: 2158–2167. PMID: 17133872.
22. Mendlová M, Šimková A. Evolution of host specificity in monogeneans parasitizing African cichlid fish. Parasit Vectors. 2014; 7: 69. doi: 10.1186/1756-3305-7-69 PMID: 24529542.
23. Brisson D, Dykhuizen DE. ospC diversity in *Borrelia burgdorferi*: different hosts are different niches. Genetics. 2004; 168: 713–722. PMID: 15514047.
24. Hanincová K, Ogden NH, Diuk-Wasser M, Pappas CJ, Lyer R, Fish D, et al. Fitness variation of *Borrelia burgdorferi* sensu stricto strains in mice. Appl Environ Microbiol. 2008; 74: 153–157. PMID: 17981941.
25. Ogden NH, Margos G, Aanensen DM, Drebot MA, Feil EJ, Hanincova K, et al. Investigation of genotypes of *Borrelia burgdorferi* in *Ixodes scapularis* ticks collected

during surveillance in Canada. Appl Environ Microbiol. 2011; 77: 3244–3254. doi: 10.1128/AEM.02636-10 PMID: 21421790.

26. Derdáková M, Dudiòák V, Brei B, Brownstein JS, Schwartz I, Fish D. Interaction and transmission of two *Borrelia burgdorferi* sensu stricto strains in a tick-rodent maintenance system. Appl Environ Microbiol. 2004; 70: 6783–6788. PMID: 15528545.
27. Ogden NH, Feil EJ, Leighton PA, Lindsay LR, Margos G, Mechai S, et al. Evolutionary aspects of emerging Lyme disease in Canada. Appl Environ Microbiol. 2015; 81: 7350–7359. doi: 10.1128/AEM.01671-15 PMID: 26296723.
28. Margos G, Tsao JI, Castillo-Ramirez S, Girard YA, Hamer SA, Hoen AG, et al. Two boundaries separate *Borrelia burgdorferi* populations in North America. Appl Environ Microbiol. 2012; 78: 6059–6067. doi: 10.1128/AEM.00231-12 PMID: 22729536.
29. Mechai S, Margos G, Feil EJ, Lindsay LR, Ogden NH. Complex population structure of *Borrelia burgdorferi* in southeastern and south central Canada as revealed by phylogeographic analysis. Appl Environ Microbiol. 2015; 81: 1309–1318. PMID: 25501480.
30. Hoen AG, Margos G, Bent SJ, Diuk-Wasser MA, Barbour A, Kurtenbach K, et al. Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. Proc Natl Acad Sci USA. 2009; 106: 15013–15018. doi: 10.1073/pnas.0903810106 PMID: 19706476.
31. Tsao JI. Reviewing molecular adaptations of Lyme *borreliosis* spirochetes in the context of reproductive fitness in natural transmission cycles. Vet Res. 2009; 40: 36..
32. Ginsberg HS, Zhioua E, Mitra S, Fischer J, Buckley PA, Verret F, et al. Woodland type and spatial distribution of nymphal *Ixodes scapularis* (*Acari: Ixodidae*). Environ Entomol. 2004; 33: 1266–1273.
33. Ogden NH, Lindsay LR, Hanincova K, Barker IK, Bigras-Poulin M, Charron DF, et al. Role of migratory birds in introduction and range expansion of Ixodes scapularis ticks and of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. Appl Environ Microbiol. 2008; 74: 1780–1790. doi: 10.1128/AEM.01982-07 PMID: 18245258.
34. Ogden NH, Radojevic M,Wu X, Duvvuri VR, Leighton PA, Wu J. Estimated effects of projected climate change on the basic reproductive number of the Lyme disease vector

Ixodes scapularis. Environ Health Perspect. 2014; 122: 631–638. doi: 10.1289/ehp.1307799 PMID: 24627295.

35. Leighton PA, Koffi JK, Pelcat Y, Lindsay LR, Ogden NH. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. J Appl Ecol. 2012; 49: 457–464.
36. Ogden NH, Mechai S, Margos G. Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. Front Cell Infect Microbiol. 2013; 29: 46.
37. Hamer SA, Tsao JI, Walker ED, Hickling GJ. Invasion of the Lyme disease vector *Ixodes scapularis*: Implications for *Borrelia burgdorferi* endemicity. Ecohealth. 2010; 7: 47–63. doi: 10.1007/s10393-010-0287-0 PMID: 20229127.
38. Ogden NH, Bouchard C, Kurtenbach K, Margos G, Lindsay LR, Trudel L, et al. Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of Lyme disease risk emergence in Canada. Environ Health Perspect. 2010; 118: 909–914. doi: 10.1289/ehp.0901766PMID: 20421192.
39. Bouchard C, Beauchamp G, Nguon S, Trudel L, Milord F, Lindsay LR, et al. Associations between *Ixodes scapularis* ticks and small mammal hosts in a newly-endemic zone in southeastern Canada: implications for *Borrelia burgdorferi* transmission. Ticks Tick-Borne Dis. 2011; 2: 183–90. doi: 10.1016/j.ttbdis.2011.03.005 PMID: 22108010.
40. Margos G, Gatewood AG, Aanensen DM, Hanincova K, Terekhova D, Vollmer SA, et al. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. Proc Natl Acad Sci USA. 2008; 105: 8730–8735. doi: 10.1073/pnas.0800323105PMID: 18574151.
41. Bunikis J, Garpmo U, Tsao J, Berglund J, Fish D, Barbour AG. Sequence typing reveals extensive strain diversity of the Lyme *borreliosis* agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. Microbiology. 2004; 150: 1741–1755. PMID: 15184561.
42. Qiu WG, Dykhuizen DE, Acosta MS, Luft BJ. Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. Genetics. 2002; 160: 833–849. PMID: 11901105.

43. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol. 2004; 186: 1518–1530. PMID: 14973027.
44. Teixeira AS, Monteiro PT, Carriço JA, Ramirez M, Francisco AP. Not Seeing the Forest for the Trees: Size of the Minimum Spanning Trees (MSTs) Forest and Branch Significance in MST-Based Phylogenetic Analysis. PLoS ONE. 2015; 10: e0119315. doi: 10.1371/journal.pone.0119315 PMID: 25799056.
45. Ronquist F, Teslenko M, van der Mark P, Ayres DL, Darling A, Höhna S, et al. MrBayes 3.2: efficient Bayesian phylogenetic inference and model choice across a large model space. Syst Biol. 2012; 61: 539–542. doi: 10.1093/sysbio/sys029 PMID: 22357727.
46. Francisco AP, Bugalho M, Ramirez M, Carrico JA. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinformatics. 2009; 10: 152. doi: 10.1186/1471-2105-10-152 PMID: 19450271.
47. Liveris D, Gazumyan A, Schwartz I. Molecular typing of *Borrelia burgdorferi* sensu lato by PCR-restriction fragment length polymorphism analysis. J Clin Microbiol. 1995; 33: 589–595. PMID: 7751362.
48. Hanincova K, Liveris D, Sandigursky S, Wormser GP, Schwartz I. Borrelia burgdorferi sensu stricto is clonal in patients with early Lyme *borreliosis*. Appl Environ Microbiol. 2008; 74: 5008–5014. doi: 10.1128/AEM.00479-08 PMID: 18539816.
49. Li W, Cowley A, Uludag M, Gur T, McWilliam H, Squizzato S, et al. The EMBL-EBI bioinformatics web and programmatic tools framework. Nucl Acids Res. 2015; 43: 580–584..
50. Gotelli NJ, Colwell RK. Quantifying biodiversity: procedures and pitfalls in the measurement and comparison of species richness. Ecol Lett. 2001; 4: 379–391.
51. Gotelli NJ, Colwell RK. Estimating species richness. In: Magurran AE, McGill BJ. Biological diversity: frontiers in measurement and assessment. New York: Oxford University Press; 2010. pp. 39–54.
52. Sanders HL. Marine benthic diversity: A comparative study. Am Nat. 1968; 102: 243–282.
53. Heck KL Jr, Van Belle G, Simberloff D. Explicit calculation of the rarefaction diversity measurement and the determination of sufficient sample size. Ecology. 1975; 56: 1459–14561.

54. Colwell RK, Coddington JA. Estimating terrestrial biodiversity through extrapolation. Philos Trans R Soc Lond B Biol Sci. 1994; 345: 101–118. PMID: 7972351
55. Hammer Ø, Harper DAT, Ryan PD. PAST: Paleontological statistics software package for education and data analysis. Palaeontologia Electronica. 2001; 4: 9.
56. Margos G, Castillo-Ramirez S, Hoen AG. Phylogeography of Lyme borreliosis-group spirochetes and methicillin-resistant *Staphylococcus aureus*. Parasitology. 2012; 139: 1952–1965. doi: 10.1017/S0031182012000741 PMID: 22617338.
57. Harville DA. Maximum likelihood approaches to variance component estimation and to related problems. J Amer Statist Assoc. 1977; 72: 320–338.
58. Zhu M. Analyzing Multilevel models with the GLIMMIX procedure. SAS Institute Inc. SAS026-2014.
59. Gatewood AG, Liebman KA, Vourc'h G, Bunikis J, Hamer SA, Cortinas R, et al. Climate and tick seasonality are predictors of *Borrelia burgdorferi* genotype distribution. Appl Environ Microbiol. 2009; 75: 2476–2483. doi: 10.1128/AEM.02633-08 PMID: 19251900
60. Ogden NH, Bigras-Poulin M, O'Callaghan CJ, Barker IK, Lindsay LR, Maarouf A, et al. Tick seasonality, host infection dynamics and fitness of *Ixodes scapularis*-borne pathogens. Parasitology. 2007; 134: 209–227.
61. Tyron CA, Snyder DP. Biology of the eastern chipmunk, (*Tamias striatus*): life tables, age, distributions, and trends in population numbers. J Mammal. 1973; 54: 145–168.
62. Schug MD, Vessey SH, Korytko AI. Longevity and survival in a population of white-footed mice (*Peromyscus leucopus*). J Mammal. 1991; 72: 360–366
63. Devevey G, Dang T, Graves CJ, Murray S, Brisson D. First arrived takes all: inhibitory priority effects dominate competition between co-infecting *Borrelia burgdorferi* genotypes. BMC Microbiol. 2015; 15: 61. doi: 10.1186/s12866-015-0381-0 PMID: 25887119.
64. Wang IN, Dykhuizen DE, Qiu W, Dunn JJ, Bosler EM, Luft BJ. Genetic diversity of ospC in a local population of *Borrelia burgdorferi* sensu stricto. Genetics. 1999; 151: 15–30. PMID: 9872945.
65. Strandh M, Raberg L. Within-host competition between *Borrelia afzelii* ospC genotypes in wild hosts as revealed by massively parallel amplicon sequencing. Philos Trans R Soc Lond B Biol Sci. 2015; 370:1675.

66. Voordouw MJ, Lachish S, Dolan MC. The Lyme disease pathogen has no effect on the survival of its rodent reservoir host. PloS One 2015; 10: 2.
67. Margos G, Hojgaard A, Lane RS, Cornet M, Fingerle V, Rudenko N, et al. Multilocus sequence analysis of *Borrelia bissettii* strains from North America reveals a new *Borrelia* species, Borrelia kurtenbachii. Ticks Tick Borne Dis. 2010; 1: 151–158. doi: 10.1016/j.ttbdis.2010.09.002 PMID: 21157575.
68. Anderson JF, Johnson RC, Magnarelli LA, Hyde FW. Identification of endemic foci of Lyme disease: isolation of *Borrelia burgdorferi* from feral rodents and ticks (*Dermacentor variabilis*). J Clin Microbiol. 1985; 22: 36–38. PMID: 3926816.
69. Mannelli A, Kitron U, Jones CJ, Slajchert TL. Role of the eastern chipmunk as a host for immature *Ixodes dammini* (*Acari: Ixodidae*) in northwestern Illinois. J Med Entomol. 1993; 30: 87–93. PMID: 8433349.
70. McLean RG, Ubico SR, Cooksey LM. Experimental infection of the eastern chipmunk (*Tamias striatus*) with the Lyme disease spirochete (*Borrelia burgdorferi*). J Wildl Dis. 1993; 29: 527–532. PMID: 8258849.
71. Slajchert T, Kitron UD, Jones CJ, Mannelli A. Role of the eastern chipmunk (Tamias striatus) in the epizootiology of Lyme *borreliosis* in northwestern Illinois, USA. J Wildl Dis. 1997; 33: 40–46. PMID: 9027689.
72. Marsot M, Sigaud M, Chapuis JL, Ferquel E, Cornet M, Vourc'h G. Introduced Siberian chipmunks (Tamias sibiricus barberi) harbor more-diverse *Borrelia burgdorferi* sensu lato genospecies than native bank voles (*Myodes glareolus*). Appl Environ Microbiol. 2011; 77: 5716–5721. doi: 10.1128/AEM. 01846-10 PMID: 21705536.
73. Marsot M, Chapuis JL, Gasqui P, Dozieres A, Masseglia S, Pisanu B, et al. Introduced Siberian chipmunks (*Tamias sibiricus barberi*) contribute more to Lyme borreliosis risk than native reservoir rodents. PLoS One. 2013; 8: e55377. doi: 10.1371/journal.pone.0055377 PMID: 23383170.
74. Barbour AG, Bunikis J, Fish D, Hanincova K. Association between body size and reservoir competence of mammals bearing *Borrelia burgdorferi* at an endemic site in the northeastern United States. Parasit Vectors. 2015; 8: 299. doi: 10.1186/s13071-015-0903-5 PMID: 26024881.

75. Rowe KC, Heske EJ, Brown PW, Paige KN. Surviving the ice: Northern refugia and postglacial colonization. Proc Natl Acad Sci USA. 2004; 101: 10355–10359. PMID: 15249673.
76. Kalkvik HM, Stout IJ, Doonan TJ, Parkinson CL. Investigating niche and lineage diversification in widely distributed taxa: phylogeography and ecological niche modeling of the *Peromyscus maniculatus* species group. Ecography. 2012; 35: 54–64.
77. Rowe KC, Heske EJ, Paige KN. Comparative phylogeography of eastern chipmunks and white-footed mice in relation to the individualistic nature of species. Mol Ecol. 2006; 15: 4003–4020. PMID: 17054499.
78. Price PK, Kennedy ML. Genic relationships in the white-footed mouse, *Peromyscus leucopus*, and the Cotton Mouse, *Peromyscus gossypinus*, Am Midl Nat. 1980; 103: 73–82.
79. Zheng X, Arbogast BS, Kenagy GJ. Historical demography and genetic structure of sister species: deer mice (*Peromyscus*) in the North American temperate rain forest. Mol Ecol. 2003; 12: 711–724. PMID: 12675826.
80. Dragoo JW, Lackey JA, Moore KE, Lessa EP, Cook JA, Yates TL. Phylogeography of the deer mouse (*Peromyscus maniculatus*) provides a predictive framework for research on hantaviruses. J Gen Virol. 2006; 87: 1997–2003. PMID: 16760402.
81. Simon JA, Marrotte RR, Desrosiers N, Fiset J, Gaitan J, Gonzalez A, et al. Climate change, habitat fragmentation, ticks and the white-footed mouse drive occurrence of *B. burgdorferi*, the agent of Lyme disease, at the northern limit of its distribution. Evol Appl. 2014; 7: 750–764. doi: 10.1111/eva.12165 PMID: 25469157.

Supporting Information

S1 File Supplementary Tables.

Data and statistical analyses tables.
Tables of the raw data used for the first time in the study, and of statistical analysis results.

Table A. The *rrs-rrlA* reference sequences from the study typed using the different methods in this study to identify IGS types, subtypes and RSTs (ribosomal sequence types). The level of similarity of IGS and RSP performed using multiple alignments is shown.

IGS accession No	IGS type ID[a]	IGS subtypeID[a]	IGS type ID[b]	RST ID[c]	level of similarity
AY275189	1	1A	RSP1	1	100%
AY275190	1	1A	RSP1	1	99.88%
AY275191	2	2A	RSP3	2	100%
AY275192	2	2A	RSP3	2	99.63%
AY275193	2	2A	RSP3	2	99.75%
AY275194	2	2D	RSP4	2	100%
AY275195	3	3A	RSP7	1	100%
AY275196	3	3A	RSP7	1	99.88%
AY275197	3	3A	RSP7	1	99.88%
AY275198	3	3A	RSP7	1	99.88%
AY275199	4	4A	RSP20	2	100%
AY275200	4	4A	RSP20	2	99.75%
AY275201	5	NI[d]	RSP14	3	100%
AY275202	6	6A	RSP9	3	100%
AY275203	6	6B	RSP18	3	100%
AY275204	6	6A	RSP9	3	99.88%
AY275205	7	7A	RSP10	3	100%
AY275206	7	7A	RSP10	3	99.88%
AY275207	8	8A	RSP12	3	100%
AY275208	8	8C	RSP13	3	99.88%
AY275209	8	8C	RSP13	3	100%
AY275210	8	8A	RSP12	3	99.75%
AY275211	9	NI	RSP19	3	99.75%
AY275212	5	NI	RSP14	3	99.88%

[a]*rrs-rrlA*typed according to Bunikis et *al.* (2004).
[b]*rrs-rrlA*typed according to Hanincova et *al.* (2008).
[c]*rrs-rrlA*typed according to Liveris et *al.* (1995).
[d]*rrs-rrlA*typed according to Bunikis et *al.*(2004).

Table B. The dataset of 91 samples collected in localities in southern Canada and used in this study. The corresponding genotypes identified using the three typing methods (MLST, IGS, *ospC*) are shown. In this and subsequent tables DM = deer mouse, ECH = eastern chipmunk RBV = red-backed vole, RS = red squirrel and WFM = white-footed mouse. Na = sequence data not available.

NMLID	ST	ospC	IGS	RSP	RST	DNA source	Tick source	Region	Lat	Lon	Year
MB11-64	302	F	4	6	2	Tick N	DM	MB	49.05	-96.48	2011
MB11-66	46	U	8C	13	3	Tick F	Drag	MB	49.05	-96.48	2011
NW13-TI5	228	na	na	na	na	Tick M	ECH	ONRv	48.72	-94.59	2013
NW13-TI7	12	na	6A	9	3	Tick M	ECH	ONRv	48.72	-94.59	2013
NW13-TI22	234	na	7A	10	3	Tick M	ECH	ONRv	48.72	-94.59	2013
NW13-TI23	804	D	4	6	2	Tick M	ECH	ONRv	48.77	-94.66	2013
NW13-TI26	29	na	2D	4	2	Tick M	ECH	ONRv	48.72	-94.59	2013
NW13-TI31	225	F3	5	14	3	Tick M	RS	ONRv	48.72	-94.59	2013
NW13-30	222	U	4	6	2	Heart	ECH	ONRv	48.72	-94.59	2013
NW13-36	300	na	na	na	na	Heart	ECH	ONRv	48.68	-94.16	2013
NW13-43	532	J	5	16	3	Heart	ECH	ONRv	48.73	-94.61	2013
NW13-45	222	na	4	6	2	Heart	ECH	ONRv	48.72	-94.59	2013
NW13-54	32	H	2D	4	2	Heart	DM	ONRv	48.68	-94.16	2013
F09.51	519	A	4A	20	2	Tick F	Drag	QC	45.31	-73.01	2009
F09.43	1	A	1A	1	1	Tick M	Drag	QC	45.31	-73.01	2009
F09.29	16	I	7A	10	3	Tick F	Drag	QC	45.31	-73.01	2009
F09.42	16	I	7A	10	3	Tick M	Drag	QC	45.31	-73.01	2009
F09.68	8	F	4	6	2	Tick M	Drag	QC	45.31	-73.01	2009
812	9	A	4A	20	2	Tick F	Drag	QC	45.31	-73.01	2008
61	9	B	4A	20	2	Tick N	WFM	QC	45.10	-72.97	2008
130	14	G	6B	18	3	Tick N	ECH	QC	45.06	-73.28	2008
816	1	A	1A	1	1	Tick F	Drag	QC	45.31	-73.01	2008
35	3	K	2A	3	2	Tick N	WFM	QC	45.18	-73.35	2008
225	3	K	2A	3	2	Tick M	Drag	QC	45.18	-73.35	2008
42	14	G	na	na	na	Tick N	ECH	QC	45.18	-73.35	2008
154	3	K	2A	3	2	Tick N	RS	QC	45.18	-73.35	2008
85	8	F	4	6	2	Tick N	ECH	QC	45.12	-73.21	2008
620	1	A	1A	1	1	Tick L	WFM	QC	45.31	-73.01	2008
264	59	B	3	7	1	Tick F	Drag	QC	45.12	-73.21	2008
102	59	B	3	7	1	Tick L	ECH	QC	45.12	-73.21	2008
QC12a-154	4	H	2D	4	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-163	12	na	6A	9	3	Tick N	Drag	QC	45.28	-72.98	2012

QC12a-168	1	A	1	1	1	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-172	1	A	1	1	1	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-175	3	A	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-176	1	A	1	1	1	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-192	1	A	1	1	1	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-198	12	na	na	na	na	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-206	3	na	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-207	3	na	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-212	315	na	5	14	3	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-218	1	na	1	1	1	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-222	806	na	6A	9	3	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-228	3	k	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-249	19	E	9	19	3	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-251	3	k	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-254	3	k	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-258	807	D	5	16	3	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-266	3	K	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-275	1	na	1	1	1	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-280	36	na	na	na	na	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-285	12	na	6A	9	3	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-286	16	na	7A	10	3	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-293	3	k	2A	3	2	Tick N	Drag	QC	45.28	-72.98	2012
QC12a-294	12	M	6A	9	3	Tick N	Drag	QC	45.28	-72.98	2012
BH.2.7	9	na	4A	20	2	Tick F	Drag	MR	44.37	-64.27	2006
L.37	535	H	6B	18	3	Tick L	WFM	MR	44.37	-64.27	2006
L.50	9	N	4A	20	2	Tick L	WFM	MR	44.53	-64.50	2006
BH.2.66	9	N	na	na	na	Tick M	Drag	MR	44.37	-64.27	2006
L.28	536	K	2A	3	2	Tick L	WFM	MR	44.53	-64.50	2006
BH.2.59	3	K	na	na	na	Tick M	Drag	MR	44.37	-64.27	2006
BH.2.55	1	A	na	na	na	Tick M	Drag	MR	44.37	-64.27	2006
BH.2.51	537	B	na	na	na	Tick F	Drag	MR	44.37	-64.27	2006
BH.2.50	16	I	7A	10	3	Tick F	Drag	MR	44.37	-64.27	2006
BH.2.35	3	K	2A	3	2	Tick M	Drag	MR	44.37	-64.27	2006
BH.2.58	59	B	na	na	na	Tick M	Drag	MR	44.37	-64.27	2006
BH.2.29	12	I	6A	9	3	Tick M	Drag	MR	44.37	-64.27	2006
SR.1.2	12	M	na	na	na	Tick M	Drag	MR	44.55	-64.58	2006
L.38	538	K	2A	3	2	Tick L	WFM	MR	44.53	-64.50	2006
ER.13	12	M	6A	9	3	Tick N	Drag	MR	43.77	-65.23	2006
CEM.3	19	E	9	19	3	Tick N	Drag	MR	44.38	-64.31	2006
BH.2.27	3	K	2A	3	2	Tick M	Drag	MR	44.37	-64.27	2006
BH.2	3	K	2A	3	2	Tick M	Drag	MR	44.37	-64.27	2006
L.41	4	H	2D	4	2	Tick L	WFM	MR	44.36	-64.26	2006

L.30	1	A	1	1	1	Tick L	WFM	MR	44.37	-64.28	2006
L.12	59	B	3	7	1	Tick L	WFM	MR	44.37	-64.27	2006
CEM.1	3	K	2A	3	2	Tick M	Drag	MR	44.38	-64.31	2006
GM11-004	9	na	4A	20	2	Tick N	WFM	MR	44.70	-66.82	2011
GM11-013	36	na	4A	20	2	Tick N	WFM	MR	44.70	-66.82	2011
GM11-021	1	A	1	1	1	Tick N	WFM	MR	44.70	-66.82	2011
GM11-030	3	na	2A	3	2	Tick L	WFM	MR	44.70	-66.82	2011
GM11-039	1	na	1	1	1	Tick N	RBV	MR	44.70	-66.82	2011
Y11-053	3	K	2A	3	2	Tick L	ECH	MR	43.83	-66.12	2011
Y11-057	3	K	2A	3	2	Tick L	WFM	MR	43.83	-66.12	2011
Y11-058	3	K	2A	3	2	Tick L	WFM	MR	43.83	-66.12	2011
GC12-C31	19	E1	9	19	3	Tick L	ECH	MR	43.68	-65.33	2012
GC12-C32	14	E	6B	18	3	Tick L	ECH	MR	43.68	-65.33	2012
GC12-DM76	1	A	1	1	1	Tick L	WFM	MR	43.68	-65.33	2012
GC12-DM118	1	A	1	1	1	Tick L	WFM	MR	43.68	-65.33	2012
AC12-DM193	1	K	1	1	1	Tick L	WFM	MR	44.73	-63.67	2012
AC12-DM205	1	A	1	1	1	Tick L	WFM	MR	44.73	-63.67	2012

Table C. A contingency table of the sample sources (host and questing ticks) and the clonal complexes (CC) inferred with single locus variant criterion (SLV).

CC at SLV	Tick sources						
	DM	ECH	RBV	RSQ	WFM	Ticks	Total
CC3	0	1	0	1	4	14	20
CC4	1	0	0	0	1	1	3
CC7	0	1	0	0	1	2	4
CC8	0	1	0	0	0	1	2
CC12	0	1	0	0	0	8	9
CC16	0	0	0	0	0	4	4
CC19	0	1	0	0	0	2	3
CC29	0	1	0	0	0	0	1
CC34	0	4	0	0	0	1	5
CC36	0	0	0	0	4	5	9
CC37	0	0	0	0	0	1	1
CC48	0	1	0	0	0	0	1
CC222	1	2	0	0	0	0	3
ST225	0	0	0	1	0	0	1
CC228	0	1	0	0	0	0	1
ST234	0	1	0	0	0	0	1
CC403	0	0	1	0	7	9	17
ST519	0	0	0	0	0	1	1
ST535	0	0	0	0	1	0	1
ST536	0	0	0	0	1	0	1
ST538	0	0	0	0	1	0	1
ST641	0	1	0	0	0	0	1
ST644	0	0	0	0	0	1	1
Total	2	16	1	2	20	50	91

Table D. The contingency table of the sample sources (host and questing ticks) and the CCs inferred at double locus variant (DLV).

CC at DLV	Tick sources						
	DM	ECH	RBV	RSQ	WFM	Ticks	Total
CC3	1	2	1	1	13	24	42
CC7	0	1	0	0	1	2	4
CC12	0	1	0	0	0	8	9
CC16	0	0	0	0	0	4	4
CC19	0	1	0	0	0	2	3
CC29	0	1	0	0	0	0	1
CC34	0	5	0	0	0	1	6
CC36	1	2	0	0	4	7	14
CC37	0	0	0	0	0	1	1
ST225	0	0	0	0	0	1	1
CC228	0	0	0	1	0	0	1
ST234	0	1	0	0	0	0	1
ST535	0	1	0	0	0	0	1
ST536	0	0	0	0	1	0	1
CC641	0	1	0	0	0	0	1
ST644	0	0	0	0	0	1	1
Total	2	16	1	2	20	50	91

Table E. Summary of the Chi-Square Goodness of Fit Test of the correspondence analysis of the CC at SLV

Dimension					Proportion of Inertia		Confidence Singular Value	
	Singular Value	Inertia	Chi Square	Sig.	Accounted for	Cumulative	Standard Deviation	Correlation
1	0.805	0.648			0.373	0.373	0.058	0.065
2	0.715	0.511			0.294	0.667	0.223	
3	0.564	0.318			0.183	0.849		
4	0.474	0.224			0.129	0.978		
5	0.194	0.038			0.022	1.000		
Total		1.739	158.281	0.002	1.000	1.000		

Table F. Overview of scores of the different sample sources (host species and questing ticks) in the correspondence analysis of the CC at SLV

host	Score in Dimension				Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
	Mass	1	2	Inertia	1	2	1	2	Total
DM	0.022	1.424	-0.077	0.311	0.055	0.000	0.115	0.000	0.116
ECH	0.176	1.822	0.190	0.498	0.725	0.009	0.944	0.009	0.953
RBV	0.011	-0.732	-0.457	0.048	0.007	0.003	0.099	0.034	0.134
RS	0.022	-1.083	5.532	0.503	0.032	0.941	0.041	0.956	0.997
WFM	0.220	-0.618	-0.320	0.232	0.104	0.032	0.291	0.069	0.360
Ticks	0.549	-0.335	-0.142	0.147	0.076	0.015	0.337	0.054	0.391
Total	1.000			1.739	1.000	1.000			

Table G. Overview of scores of the different CCs inferred at SLV (at SLV) and singleton STs

CC at SLV		Score in Dimension			Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
	Mass	1	2	Inertia	1	2	1	2	Total
CC3	0.22	-0.399	0.172	0.044	0.043	0.009	0.634	0.105	0.739
CC4	0.033	0.195	-0.251	0.157	0.002	0.003	0.006	0.009	0.016
CC7	0.044	0.166	-0.145	0.004	0.002	0.001	0.234	0.158	0.391
CC8	0.022	0.923	0.034	0.019	0.023	0	0.783	0.001	0.784
CC12	0.099	-0.118	-0.147	0.05	0.002	0.003	0.022	0.03	0.052
CC16	0.044	-0.416	-0.198	0.036	0.009	0.002	0.17	0.034	0.204
CC19	0.033	0.477	-0.044	0.015	0.009	0	0.416	0.003	0.419
CC29	0.011	2.262	0.266	0.052	0.07	0.001	0.879	0.011	0.89
CC34	0.055	1.727	0.173	0.149	0.204	0.002	0.885	0.008	0.893
CC36	0.099	-0.572	-0.309	0.046	0.04	0.013	0.572	0.148	0.721
CC37	0.011	-0.416	-0.198	0.009	0.002	0.001	0.17	0.034	0.204
CC48	0.011	2.262	0.266	0.052	0.07	0.001	0.879	0.011	0.89
CC222	0.033	2.098	0.141	0.217	0.18	0.001	0.538	0.002	0.54
ST225	0.011	-1.345	7.738	0.489	0.025	0.92	0.033	0.962	0.995
CC228	0.011	2.262	0.266	0.052	0.07	0.001	0.879	0.011	0.89
ST234	0.011	2.262	0.266	0.052	0.07	0.001	0.879	0.011	0.89
CC403	0.187	-0.59	-0.327	0.111	0.081	0.028	0.469	0.128	0.598
ST519	0.011	-0.416	-0.198	0.009	0.002	0.001	0.17	0.034	0.204
ST535	0.011	-0.768	-0.448	0.039	0.008	0.003	0.134	0.04	0.174
ST536	0.011	-0.768	-0.448	0.039	0.008	0.003	0.134	0.04	0.174
ST538	0.011	-0.768	-0.448	0.039	0.008	0.003	0.134	0.04	0.174
ST641	0.011	2.262	0.266	0.052	0.07	0.001	0.879	0.011	0.89
ST644	0.011	-0.416	-0.198	0.009	0.002	0.001	0.17	0.034	0.204
Total	1.00			1.739	1.00	1.00			

Table H. Summary of the Chi-Square Goodness of Fit Test of the correspondence analysis of the CC at DLV

Dimension	Singular Value	Inertia	Chi Square	Sig.	Proportion of Inertia		Confidence Singular Value	
					Accounted for	Cumulative	Standard Deviation	Correlation
1	.722	.521			.439	.439	.142	.965
2	.683	.467			.393	.832	.119	
3	.412	.170			.143	.975		
4	.146	.021			.018	.993		
5	.093	.009			.007	1.000		
Total		1.188	108.136	.007	1.000	1.000		

Table I. Overview of scores of the different CCs (at DLV) and singleton STs

CC at DLV	Score in Dimension				Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
	Mass	1	2	Inertia	1	2	1	2	Total
CC3	0.462	-0.267	0.325	0.068	0.046	0.071	0.352	0.493	0.845
CC7	0.044	0.267	-0.063	0.004	0.004	0.000	0.541	0.028	0.570
CC12	0.099	0.073	0.238	0.050	0.001	0.008	0.008	0.076	0.084
CC16	0.044	-0.127	0.513	0.036	0.001	0.017	0.014	0.219	0.234
CC19	0.033	0.473	-0.312	0.015	0.010	0.005	0.366	0.150	0.516
CC29	0.011	1.672	-1.961	0.052	0.043	0.062	0.430	0.560	0.991
CC34	0.066	1.372	-1.548	0.198	0.172	0.231	0.453	0.546	0.999
CC36	0.154	0.061	0.213	0.027	0.001	0.010	0.015	0.177	0.193
CC37	0.011	-0.127	0.513	0.009	0.000	0.004	0.014	0.219	0.234
ST225	0.011	-6.233	-4.900	0.489	0.591	0.386	0.630	0.369	0.999
CC228	0.011	1.672	-1.961	0.052	0.043	0.062	0.430	0.560	0.991
ST234	0.011	1.672	-1.961	0.052	0.043	0.062	0.430	0.560	0.991
ST535	0.011	-0.351	0.683	0.039	0.002	0.008	0.025	0.090	0.115
ST536	0.011	-0.351	0.683	0.039	0.002	0.008	0.025	0.090	0.115
CC641	0.011	1.672	-1.961	0.052	0.043	0.062	0.430	0.560	0.991
ST644	0.011	-0.127	0.513	0.009	0.000	0.004	0.014	0.219	0.234
Total	1.000			1.188	1.000	1.000			

Table J. Overview of scores of the different sample sources (host species and questing ticks) with CCs inferred at DLV.

Host	Score in Dimension				Contribution				
					Of Point to Inertia of Dimension		Of Dimension to Inertia of Point		
	Mass	1	2	Inertia	1	2	1	2	Total
DM	0.022	-0.143	0.394	0.026	0.001	0.005	0.013	0.091	0.103
ECH	0.176	1.207	-1.340	0.402	0.355	0.462	0.460	0.537	0.997
RBV	0.011	-0.370	0.476	0.013	0.002	0.004	0.085	0.133	0.217
RS	0.022	-4.501	-3.348	0.490	0.617	0.361	0.656	0.344	1.000
WFM	0.220	-0.254	0.467	0.151	0.020	0.070	0.068	0.217	0.284
Ticks	0.549	-0.092	0.351	0.107	0.006	0.099	0.031	0.431	0.462
Total	1.000			1.188	1.000	1.000			

Table K. Associations of host species with STs of CC34 at SLV. The models did not include of visits as a random effect because this was not significant, but did include site ID nested by region as a random effect because this was significant ($\beta = 0.03 \pm 0.01$; Wald-test = 1.8; P = 0.0358 corresponding to 49.6 % of the total variation). Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance ($\chi^2 = 4.18$, df = 3, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	P > F
Host	5	63	4.28	0.002

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.01320	0.07624	22	0.17	0.864
DM	0.07628	0.1655	63	0.46	0.646
ECH	0.3340	0.09421	63	3.55	0.001
RBV	-0.0237	0.2311	63	-0.10	0.918
RS	0.06330	0.1556	63	0.41	0.685
WFM	-0.0344	0.07780	63	-0.44	0.659
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.01298	0.07381	22	0.18	0.862
ECH	0.3207	0.09097	66	3.53	<0.001
Other host spp.	-0.0146	0.07325	66	-0.20	0.843
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0	0.05547	39	0.00	1.0
ECH	0.2500	0.089	39	2.82	0.008
Other host spp.: reference					

Table L. Associations of host species with STs of CC34 at DLV. The models did not include of visits as a random effect because this was not significant, but did include site ID nested by region as a random effect because this was significant ($\beta = 0.04 \pm 0.02$; Wald-test = 2.07; P = 0.0192 corresponding to 55.4 % of the total variation). Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance ($\chi^2 = 3.06$, df = 3, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	$P > F$
Host	5	63	7.00	<0.001

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.01873	0.08054	22	0.23	0.818
DM	-0.1672	0.1689	63	-0.99	0.326
ECH	0.4267	0.09729	63	4.39	<0.001
RBV	-0.0247	0.2366	63	-0.10	0.917
RS	0.1187	0.1576	63	0.75	0.454
WFM	-0.0296	0.07999	63	-0.37	0.713
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.01951	0.07786	22	0.25	0.805
ECH	0.4083	0.09428	66	4.33	<0.001
Other host spp.	-0.0287	0.07564	66	-0.38	0.705
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0	0.05938	39	0.00	1.0
ECH	0.3125	0.09505	39	3.29	0.002
Other host spp.: reference					

Table M. Associations of host species with STs of CC403 at SLV. The random effects were not significant and not included in the models. Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance (χ^2 = 5.22, df = 3, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	5	85	2.70	0.026

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.1800	0.05298	85	3.40	0.001
DM	-0.1800	0.2702	85	-0.67	0.507
ECH	-0.1800	0.1076	85	-1.67	0.098
RBV	0.8200	0.3784	85	2.17	0.033
RS	-0.1800	0.2702	85	-0.67	0.507
WFM	0.3500	0.08377	85	4.18	<0.001
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.1800	0.05411	88	3.33	0.001
WFM	0.3500	0.08555	88	4.09	<0.001
Other host spp.	0.04762	0.08349	88	0.57	0.570
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.04762	0.08197	39	0.58	0.565
WFM	0.3024	0.1174	39	2.58	0.014
Other host spp.: reference					

Table N. Associations of host species with STs of CC4 at SLV. The models did not include visits as a random effect because this was not significant, but did include site ID nested by region as a random effect because this was significant (β = 0.027 ± 0.014; Wald-test = 1.9; *P* = 0.0312 corresponding to 58.2 % of the total variation). Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance (χ^2 = 6.97, df = 3, P >0.05).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	5	63	3.21	0.012

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.06703	0.06409	22	1.05	0.307
DM	0.4501	0.1321	63	3.41	0.001
ECH	-0.0751	0.07650	63	-0.98	0.329
RBV	-0.0323	0.1854	63	-0.17	0.862
RS	-0.0557	0.1228	63	-0.45	0.652
WFM	-0.0074	0.06277	63	-0.12	0.907
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.06157	0.06311	22	0.98	0.340
DM	0.4662	0.1296	66	3.60	<0.001
Other host spp.	-0.0279	0.05841	66	-0.48	0.633
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.02564	0.03113	39	0.82	0.564
DM	0.4744	0.1410	39	3.37	0.014
Other host spp.: reference					

Table O. Associations of host species with RST1 type IGS sequences. The models did not include of visits as a random effect because this was not significant, but did include site ID nested by region as a random effect because this was significant (β = 0.026 ± 0.02; Wald-test = 1.07; *P* = 0.1426). Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance (χ^2 = 5.01, df = 3, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	5	74	2.38	0.046

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.1905	0.05956	74	3.20	0.002
DM	-0.1905	0.2794	74	-0.68	0.497
ECH	-0.1905	0.1225	74	-1.55	0.124
RBV	0.8095	0.3906	74	2.07	0.042
RS	-0.1905	0.2794	74	-0.68	0.497
WFM	0.3500	0.08631	74	4.05	<0.001
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.1905	0.06084	77	3.13	0.002
WFM	0.3500	0.08817	77	3.97	<0.001
Other host spp.	0.05556	0.09293	77	0.60	0.552
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.05556	0.09208	36	0.60	0.550
WFM	0.2944	0.1269	36	2.32	0.026
Other host spp.: reference					

Table P. Associations of host species with RST2 type IGS sequences (IGS4). The random effects were not significant and not included in the model. Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance ($\chi^2 = 3.64$, df = 3, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	5	74	4.55	0.001

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.02381	0.03696	74	0.64	0.521
DM	0.4762	0.1734	74	2.75	0.007
ECH	0.2839	0.07602	74	3.73	<0.001
RBV	-0.0238	0.2424	74	-0.10	0.922
RS	-0.0238	0.1734	74	-0.14	0.891
WFM	-0.0238	0.06507	74	-0.37	0.715
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.02381	0.03814	77	0.62	0.534
ECH	0.2839	0.07846	77	3.62	<0.001
Other host spp.	0.01619	0.06244	77	0.26	0.796
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.04000	0.06437	36	0.62	0.538
ECH	0.2677	0.1101	36	2.43	0.020
Other host spp.: reference					

Table Q. Associations of host species with *ospC* G. The random effect of number of visits was significant and included in the model (β = 0.025 ± 0.009; Wald-test = 2.76; *P* = 0.0029 corresponding to 74.6 % of the total variation), but the random effect of site ID nested by region was not significant and not included in the model. Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance (χ^2 = 2.88, df = 2, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	4	33	5.67	0.001

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	-0.00902	0.05113	31	-0.18	0.861
DM	0.009021	0.1449	33	0.06	0.951
ECH	0.2716	0.07427	33	3.66	<0.001
RS	0.1447	0.1207	33	1.20	0.239
WFM	-0.02032	0.05768	33	-0.35	0.727
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	-0.00216	0.04874	32	-0.04	0.965
ECH	0.2471	0.07165	34	3.45	0.002
Other host spp.	-0.00637	0.05550	34	-0.11	0.909
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	-0.00161	0.05148	5	-0.03	0.976
ECH	0.2113	0.07590	5	2.78	0.039
Other host spp.: reference					

Table R. Associations of host species with *ospC* A. The random effects were not significant and not included in the model. Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance ($\chi^2 = 0.88$, df = 2, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	4	64	4.73	0.002

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.2632	0.06802	64	3.87	0.003
DM	0.01503	0.2942	64	0.05	0.960
ECH	-0.0189	0.1345	64	-0.14	0.889
RS	0.02967	0.2910	64	0.10	0.919
WFM	0.3525	0.1040	64	3.39	0.002
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.2632	0.06698	66	3.93	0.002
WFM	0.3529	0.1001	66	3.52	0.008
Other host spp.	-0.2632	0.1291	66	-2.04	0.050
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.00002	0.09779	29	0.00	1.0
WFM	0.3529	0.1321	29	2.67	0.012
Other host spp.: reference					

Table S. Associations of host species with *ospC* H. The random effects were not significant and not included in the model. Reducing the parameters from the full model to the minimal model including questing ticks did not significantly affect model deviance ($\chi^2 = 1.19$, df = 2, P >0.1).

Type III Tests of Fixed Effects

Factor	Num DF	Den DF	F Value	*P* > F
Host	4	64	2.62	0.043

Full model

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.02632	0.03649	64	0.72	0.473
DM	0.4737	0.1632	64	2.90	0.005
ECH	-0.0263	0.07995	64	-0.33	0.743
RS	-0.0263	0.1632	64	-0.16	0.872
WFM	0.09133	0.06564	64	1.39	0.167
Questing ticks: reference					

Minimal model with questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.02632	0.03647	66	0.72	0.473
DM	0.4737	0.1631	66	2.90	0.005
Other host spp.	0.04265	0.05543	66	0.77	0.444
Questing ticks: reference					

Minimal model without questing ticks

Factor	Estimate	Standard Error	DF	t Value	P > |t|
Intercept	0.06897	0.05300	29	1.30	0.203
DM	0.4310	0.2086	29	2.07	0.048
Other host spp.: reference					

Table T. Comparisons of the minimal models (with questing ticks) presented in Tables S4a to S4i and the respective intercept-only models.

Outcome	Model	-2 Log Likelihood	DF	chisq	*P* > |chisq|
CC34S	Minimal Model	8.67	1	12.58	<0.001
	Intercept Model	21.25			
CC34D	Minimal Model	10.06	1	6.42	0.011
	Intercept Model	3.64			
CC403S	Minimal Model	86.76	2	6.42	0.040
	Intercept Model	80.34			
CC4S	Minimal Model	49.21	1	12.44	<0.001
	Intercept Model	61.65			
RST1	Minimal Model	94.58	2	9.02	0.011
	Intercept Model	85.56			
RST2 (IGS4)	Minimal Model	19.78	2	6.96	0.031
	Intercept Model	12.82			
ospC G	Minimal Model	56.85	1	17.44	<0.001
	Intercept Model	74.29			
ospC A	Minimal Model	76.77	2	6.1	0.047
	Intercept Model	70.67			
ospC H	Minimal Model	0.49	1	3.94	0.047
	Intercept Model	4.43			

Table U. Statistics of the minimal models with questing tick data in Tables S4a-i when using GLMs in R software.

CC34S with region/SitesID random effect				
Factor	Estimate	Standard Error	Z Value	*P* > |z|
Intercept	-4.053	1.133	-3.578	<0.001
ECH	2.601	1.282	2.028	0.042
Other host spp.	-0.2677	< 0.0001	0.000	0.999
Questing ticks: reference				

CC34D with region/SitesID random effect				
Factor	Estimate	Standard Error	Z Value	*P* > |z|
Intercept	-4.039	1.123	-3.595	<0.001
ECH	2.895	1.252	2.312	0.021
Other host spp.	-0.3084	0.0000008	0.000	0.999
Questing ticks: reference				

CC403S				
Factor	Estimate	Standard Error	t Value	*P* > |z|
Intercept	0.18000	0.05411	3.327	0.001
WFM	0.17000	0.09949	1.679	0.097
Other host spp.	-0.13238	0.10123	-1.331	0.187
Questing ticks: reference				

CC4S with region/SitesID random effect				
Factor	Estimate	Standard Error	Z Value	*P* > |z|
Intercept	-3.8918	1.0102	-3.853	<0.001
DM	3.8918	1.7379	2.239	0.025
Other host spp.	0.2542	1.4306	0.178	0.859
Questing ticks: reference				

RST 1				
Factor	Estimate	Standard Error	t Value	*P* > |z|
Intercept	0.21429	0.06378	3.360	0.001
WFM	0.21429	0.11047	1.940	0.056

Other host spp.	-0.15546	0.11882	-1.308	0.194
Questing ticks: reference				
RST2 (IGS4)				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.02381	0.03814	0.624	0.534
ECH	0.28388	0.07846	3.61	<0.001
Other host spp.	0.01619	0.06244	0.259	0.796
Questing ticks: reference				

Table U continued.

***ospC* G with number of visits random effect**				
Factor	Estimate	Standard Error	z Value	$P > \|z\|$
Intercept	0.00002673	0.02526	0.000	1.0
ECH	0.2	0.05534	3.614	<0.001
Other host spp.	-0.0000302	0.04234	0.000	1.0
Questing ticks: reference				
***ospC* A**				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.2632	0.1291	2.039	0.045
WFM	0.3529	0.149	2.369	0.021
Other host spp.	0.00000052	0.01103	0.000	1.0
Questing ticks: reference				
***ospC* H**				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.02632	0.03647	0.722	0.473
DM	0.47368	0.16310	2.904	0.005
Other host spp.	0.04265	0.05543	0.769	0.444
Questing ticks: reference				

Table V. Statistics of the minimal models without questing tick data in Tables S4a-i when using GLMs in R software.

CC34S with region/SitesID random effect				
Factor	Estimate	Standard Error	Z Value	$P > \|z\|$
Intercept	-0.00001	0.005547	0.000	1.0
ECH	0.2500	0.008880	2.815	0.008
Other host spp. reference				
CC34D with region/SitesID random effect				
Factor	Estimate	Standard Error	Z Value	$P > \|z\|$
Intercept	-0.00001	0.005938	0.000	1.0
ECH	0.03125	0.009505	3.288	0.002
Other host spp. reference				
CC403S				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.04762	0.08197	0.581	0.565
WFM	0.30238	0.11736	2.577	0.014
Other host spp. reference				
CC4S with region/SitesID random effect				
Factor	Estimate	Standard Error	Z Value	$P > \|z\|$
Intercept	0.02564	0.03113	0.824	0.415
DM	0.47436	0.14097	3.365	0.002
Other host spp. reference				
RST 1				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.1111	0.1008	1.103	0.277
WFM	0.2889	0.1389	2.080	0.045
Other host spp. reference				

RST2 (IGS4)				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.04000	0.06437	0.621	0.538
ECH	0.26769	0.11005	2.432	0.020
Other host spp. reference				

Table V continued.

***ospC* G with number of visits random effect**				
Factor	Estimate	Standard Error	z Value	$P > \|z\|$
Intercept	0.00000003	0.005126	0.000	1.0
ECH	0.02	0.009025	2.216	0.035
Other host spp. reference				
***ospC* A**				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.00008	0.009779	0.000	1.0
WFM	0.03529	0.0 1321	2.673	0.012
Other host spp. reference				
***ospC* H**				
Factor	Estimate	Standard Error	t Value	$P > \|z\|$
Intercept	0.06897	0.05300	1.301	0.203
DM	0.43103	0.20865	2.066	0.045
Other host spp. reference				

S2 File Supplementary Figures.

These comprise an unrooted Bayesian phylogenetic tree and a minimum spanning tree of the MLST STs.

Fig. A

Unrooted Bayesian phylogenetic tree based on STs constructed without outgroup and the corresponding clonal complexes (CCs) at single locus variant (SLV) and double locus variant (DLV) performed using goeburst. The posterior probabilities produced by the tree are used to support the confidence level of each CC

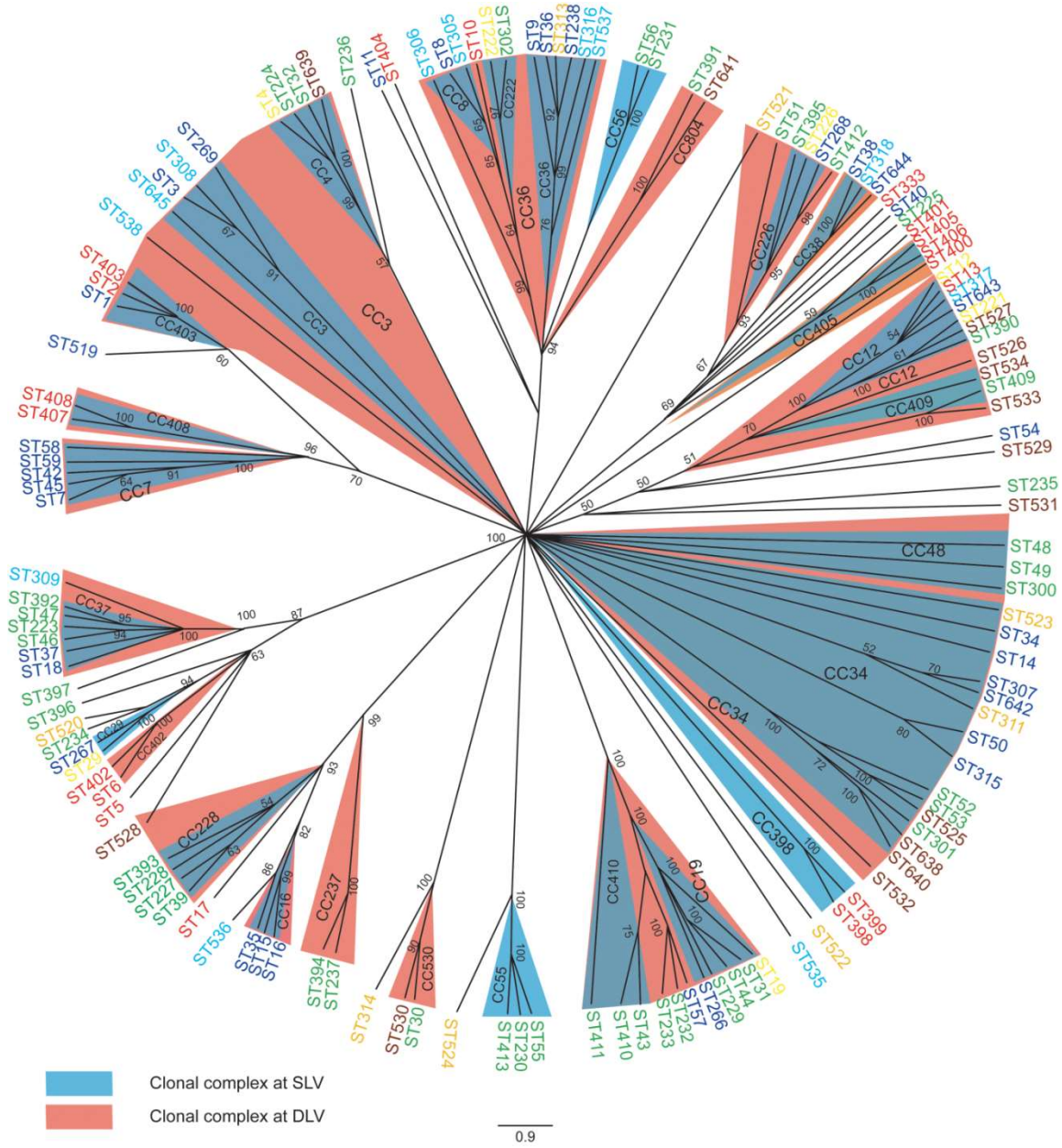
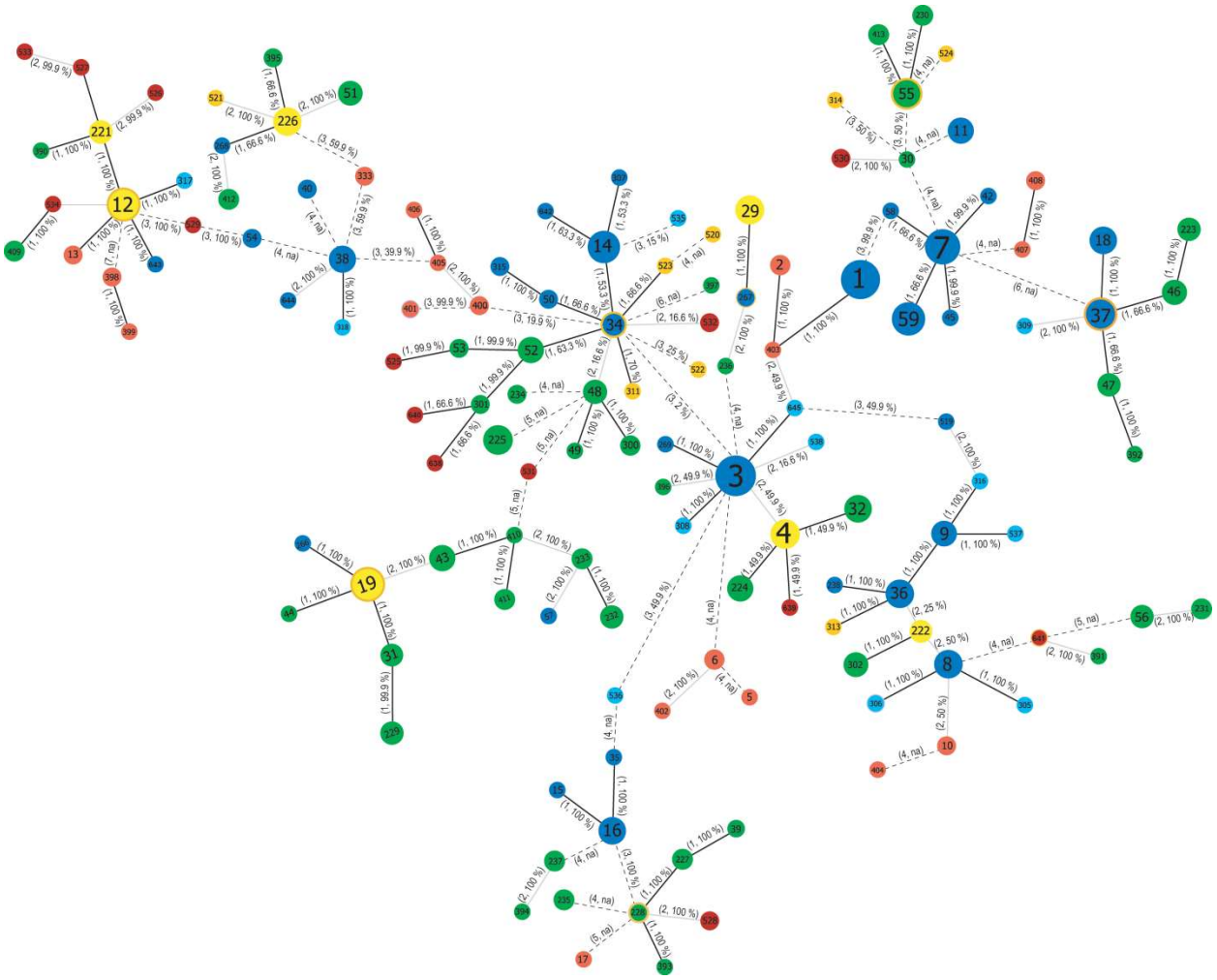


Fig. B. A minimum spanning tree of all the STs in the MLSTNet database created in GoeBURST. The STs are colour coded according to their geographic region of origin: blue, STs found in the 'northeast' (i.e. STs found in Northeastern US and also in Quebec, eastern Ontario and the Maritimes); green, STs found in the 'midwest' (i.e. STs found in Midwestern US and also in Manitoba); yellow, STs occurring in both the 'northeast' and the 'midwest'; red, STs found in California; cyan STs found only in the Maritimes; orange, STs found only at Long Point Ontario; brown, STs found only in Manitoba. Lines connecting the STs indicate the numbers of different loci between the connected STs: black lines = SLV, gray lines = DLV, dashed lines = TLV and greater. For differences > TLV the optimal phylogenetic edge produced by GoeBURST was the potential ancestry link. Statistics presented for each link are number of locus differences and the spanning edge betweenness statistic. High values at or near 100% indicate that the predicted link is unique while low values indicate that while the predicted link is the optimal one, there are many other pathways from ancestor to descendant, which may be associated with horizontal gene transfer [1]



References

Teixeira AS, Monteiro PT, Carriço JA, Ramirez M, Francisco AP. Not seeing the forest for the trees: size of the minimum spanning trees (MSTs) forest and branch significance in MST-based phylogenetic analysis. PLoS One. 2015;10(3): e0119315.

V. Chapitre III : *Utilisation des connaissances sur les souches pour mieux comprendre la dispersion de B. burgdorferi*

Evidence for an effect of landscape connectivity on *Borrelia burgdorferi* dispersion in a zone of range expansion

Short title: Landscape connectivity affects *Borrelia burgdorferi* dispersion

Samir Mechai[1,2,] Gabriele Margos[3,4,5], Edward J. Feil[6], L. Robbin Lindsay[7], Pascal Michel[8], Serge-Olivier Kotchi[2,9], Nick H. Ogden[2,9,*]

1. Département de pathologie et microbiologie, Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, J2S 7C6, Canada
2. Groupe de Recherche en Épidémiologie des Zoonoses et Santé Publique, Faculté de médecine vétérinaire, Université de Montréal, Saint-Hyacinthe, Québec, J2S 7C6, Canada
3. Ludwig Maximilians Universität München, Department for Infectious Diseases and Zoonoses, Munich, Germany
4. National Reference Centre for Borrelia, Oberschleissheim, Germany
5. Bavarian Health and Food Safety Authority, Oberschleissheim, Germany
6. Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath, United Kingdom
7. Zoonoses and Special Pathogens Division, National Microbiology Laboratory, Public Health Agency of Canada, Winnipeg, Manitoba, Canada
8. Office of the Chief Science Office, Public Health Agency of Canada, Ottawa, Ontario, K1A 0K9, Canada.
9. Public Health Risk Sciences Division, National Microbiology Laboratory, Public Health Agency of Canada, Saint-Hyacinthe, Québec, J2S 7C6, Canada.

PLOS ONE
Submitted: July 27, 2017
Status: reviewing

* Corresponding author: nicholas.ogden@phac-aspc.gc.ca

Abstract

Different strains of the Lyme disease-causing bacterium *Borrelia burgdorferi* sensu stricto cluster into phylogenetic groups that are associated with different levels of pathogenicity and, for some, specific rodent reservoir hosts. Here we explore whether landscape connectivity, by impacting host dispersal, influences *B. burgdorferi* spread patterns. This question is central to modelling spatial patterns of the spread of Lyme disease risk in the zone of northward range-expansion of *B. burgdorferi* in southeastern Canada where the study was conducted. We used multi-locus sequence typing (MLST) to characterise *B. burgdorferi* in positive ticks collected at 13 sites in southern Quebec, Canada during the early stages of *B. burgdorferi* invasion. We used mixed effects logistic regression to investigate whether landscape connectivity (probability of connectivity; PC) affected the probability that samples collected at different sites were of the same strain (MLST sequence type: ST). PC was calculated from a habitat map based on high spatial resolution (15m) Landsat 8 imagery to identify woodland habitat that are preferred by rodent hosts of *B. burgdorferi*. There was a significant positive association between the likelihood that two samples were of the same ST and PC, when PC values were grouped into three categories of low, medium and high. When analysing data for individual STs, samples at different sites were significantly more likely to be the same when PC was higher for the rodent-associated ST1. These findings support the hypothesis that dispersion trajectories of *B. burgdorferi* in general, and some rodent-associated strains in particular, are at least partly determined by landscape connectivity. This may suggest that dispersion of *B. burgdorferi* is more common by terrestrial mammal hosts (which would likely disperse according to landscape connectivity) than by birds, the dispersal of which is likely less constrained by landscape. This study suggests that accounting for landscape connectivity may improve model-based predictions of spatial spread patterns of *B. burgdorferi*. The findings are consistent with possible past dispersal patterns of *B. burgdorferi* as determined by phylogeographic studies.

Keywords: *Borrelia burgdorferi*, connectivity, landscape, Lyme disease,

Introduction

Climate change is likely to drive emergence and re-emergence of vector-borne diseases and zoonoses [1]. Public and animal health professionals are increasingly aiming to predict the patterns of emergence of these diseases, in order to assess risk and to inform disease management strategies [2]. Model-based prediction relies on knowledge of the biotic and abiotic components of the environmental niche of the pathogens and vectors, but the actual process of spread is frequently assumed or highly simplified [3]. This may be adequate for some risk assessment processes, but for the purposes of targeted surveillance, prevention and control at a local level, the mechanisms (and thus trajectories) of pathogen invasion may be crucial. Lyme disease, caused by *Borrelia burgdorferi* sensu stricto (henceforth termed *B. burgdorferi*), is a vector-borne zoonosis that is emerging in North America and causing high public health impact. *Borrelia burgdorferi* is transmitted amongst wild animal reservoir hosts, particularly birds and rodents, by the tick *Ixodes scapularis* in the upper Midwestern and northeastern USA, and southcentral and southeastern Canada [4]. The risk of Lyme disease due to *B. burgdorferi* has most recently emerged (and continues to emerge) in Canada due to northward spread of *I. scapularis* and *B. burgdorferi*, and their hosts, associated with a warming climate [5, 6, 7]. There is a critical public health need to identify other regions likely to be at risk [6]. *Borrelia burgdorferi* and immature *I. scapularis* are broadly host generalists, their hosts are woodland rodents and birds. Woodlands also provide refuges for ticks to survive over winter, and habitat for deer which are the main hosts for adult *I. scapularis* [8]. *Borrelia burgdorferi* is a diverse species with, at the time of writing, 138 strains identified to date in North America using Multi Locus Sequence Typing (MLST: [9, 10]). Recently, we found evidence for associations of different strains of *B. burgdorferi* with different rodent host species [10]. These associations do not amount to complete host specialization; rather they indicate that some strains are more efficiently transmitted by particular hosts, while the species remains a generalist [10]. Nevertheless, there is evidence that such host association may have shaped the phylogenetic tree of *B. burgdorferi*, with major clades possibly having been associated with expansions of different host species with past, glacial-interglacial climate changes [10, 11]). The evolutionary ecology of *B. burgdorferi* has public health importance as

different strains have different pathogenicity in humans [12], a phenotypic trait that may also have its origins in host associations [10].

To understand and predict the occurrence of *B. burgdorferi* and its different strains, particularly in zones of invasion such as southern Canada, we need to understand how it is dispersed. Ticks have very limited capacity for dispersal by themselves (their lateral movements being limited to a meter at best [13]), so the bacterium can be spread by either feeding ticks carried by hosts or by infected hosts themselves. Consequently, host dispersal patterns will be key to *B. burgdorferi* dispersal. There is much evidence that passerine birds migrating north in spring and south in autumn facilitate long distance dispersal of ticks and *B. burgdorferi.* This then provides a means to seed founder populations and drive range expansions [14, 15, 16]. Shorter scale dispersal by terrestrial mammals and non-migratory species may also play a role in dispersal at local spatial scales [17]. The relative contribution of birds versus mammal hosts to dispersion of *I. scapularis* and *B. burgdorferi* is unknown. However as non-woodland habitats have greater resistance to rodent dispersion than woodlands, and landscape features such as rivers and roads are significant barriers [18, 19, 20, 21], it would be expected that trajectories of dispersion by woodland rodents would be more dependent on the connectivity of woodlands than those by birds as seen in Europe [22] where *B. burgdorferi* sensu lato genospecies are strongly specialised for host species [23].

In this study we investigated the degree to which connectivity of woodland habitats influences the likelihood that *B. burgdorferi* samples collected from different field sites are of the same strain. All sites were located in a region of *I. scapularis* and *B. burgdorferi* invasion in southern Quebec. Our hypothesis was that if landscape connectivity impacts whether strains in different sites are the same, then it is likely that terrestrial hosts such as rodents are key to dispersal, and that landscape connectivity must be accounted for in modelling the spread of the pathogen at a local scale.

Material and Methods

Study area

Samples were collected in Montérégie region in southwestern Quebec, Canada (Fig 1) between 2008 and 2012. This area covers 11,851 km^2 (http://www.nrcan.gc.ca/), and is an area of Lyme disease emergence. The landscape is a patchwork of agricultural fields (60.3 %), woodlands (21.9 %), urban areas (13 %) and water bodies (3.2 %) [24]. Samples for the study were obtained from 13 well-characterised field study sites in which the habitat was deciduous forest dominated by maple trees (*Acer saccharum*: [25]), as well as five woodland sites at which ticks were also obtained from hunter killed deer (26).

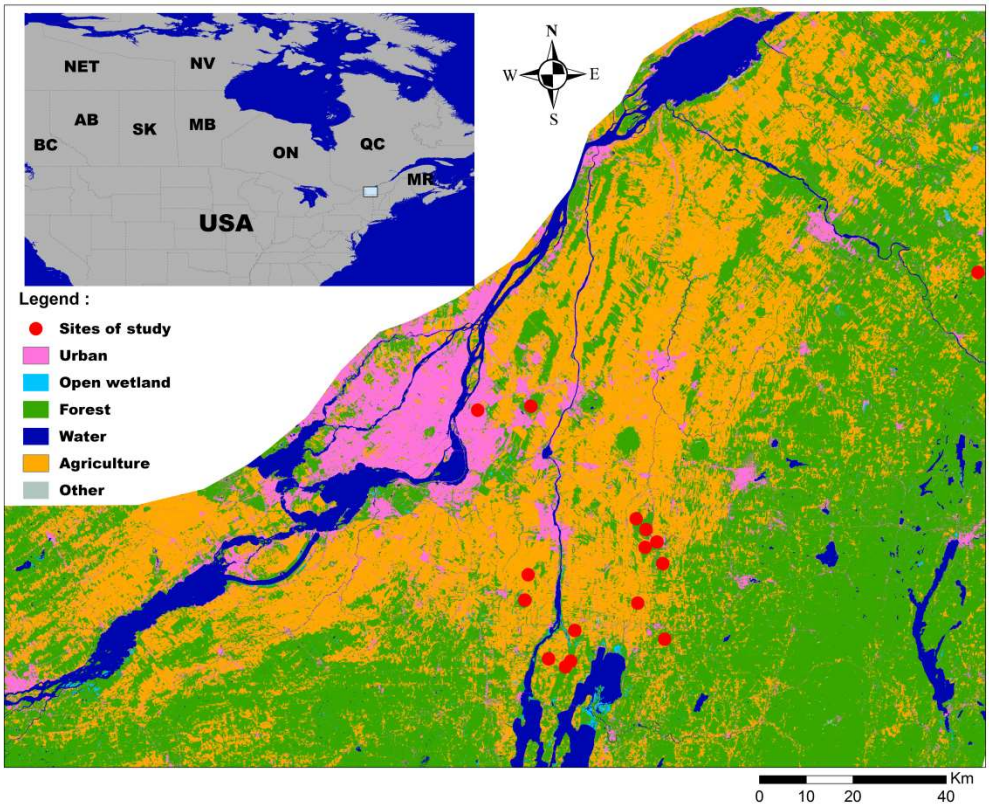


Figure 1 : Study sites and landscape classification map

The sites where ticks were collected are shown by red dots on the landscape map of southern Quebec, Canada, obtained by classifying habitat according to Landsat 8 images. Some sites that are very close are represented by one dot.

Ticks and small mammalian sampling, and B. burgdorferi genotyping

The 58 samples were all ticks that were *B. burgdorferi*-positive by polymerase chain reaction (PCR [10]). These included 39 questing ticks and 12 ticks collected from captured rodents including white footed mice (*Peromyscus leucopus* Rafinesque, 1818), eastern chipmunk (*Tamias striatus* Linnaeus, 1758) and red squirrel: (*Tamiasciurus hudsonicus* Erxleben, 1777). In addition, 7 ticks were obtained from hunter-killed white-tailed deer (*Odocoileus virginianus* Zimmermann, 1780) as previously described [26]. Samples collected at the 13 field study sites were collected by a standardized sampling regime that combined rodent trapping to collect feeding ticks and other rodent samples, and drag sampling to collect questing ticks. Some sites were sampled more than once during the 2008-2012 period. In the case of tick samples from rodents and deer, only one tick was used in analyses.

Borrelia burgdorferi in the samples had been genotyped by MLST as previously described [27]. Briefly, after DNA extraction, 8 housekeeping genes (*clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB* and *uvrA*) were amplified and sequenced [9]. In the MLST scheme (http://pubmlst.org/borrelia/) a numerical ID is assigned for each different allele sequence, and by aligning the 8 genes an allelic profile is created for each sample. The allelic profile over all genes defines the sequence type (ST), which is the definition of 'strain' used in this study. The MLST data used in this study comes from 58 DNA samples (comprising 15 STs; Table 1) that have been published previously [10, 27] and all are available at http://www.pubmlst.org/borrelia/. As previously described [27] ticks carrying possible mixed strain infections were eliminated from analysis by examination of MLST sequences and the more variable *rrs-rrlA* (16S-23S) intergenic spacer (IGS) region.

Table 1: The dataset of 58 samples collected in localities in southern Quebec.

Sample ID	ST	CC	Tick	Tick source	Year	Latitude	Longitude
F09.51	519	ST519	F	Drag	2009	45.31	-73.01
F09.43	1	CC403	M	Drag	2009	45.31	-73.01
F09.29	16	CC16	F	Drag	2009	45.31	-73.01
F09.42	16	CC16	M	Drag	2009	45.31	-73.01
F09.68	8	CC8	M	Drag	2009	45.31	-73.01
812	9	CC36	F	Drag	2008	45.31	-73.01
61	9	CC36	N	WFM	2008	45.10	-72.97
25.1	36	CC36	F	WTD	2008	45.06	-73.22
130	14	CC34	N	ECH	2008	45.06	-73.28
816	1	CC403	F	Drag	2008	45.31	-73.01
60.2	1	CC403	F	WTD	2008	45.27	-73.02
35	3	CC3	N	WFM	2008	45.18	-73.35
225	3	CC3	M	Drag	2008	45.18	-73.35
789	14	CC34	F	Drag	2008	45.54	-73.46
28.4	1	CC403	F	WTD	2008	45.05	-73.24
129	14	CC34	N	ECH	2008	45.06	-73.28
127	14	CC34	N	ECH	2008	45.06	-73.28
802	3	CC3	F	Drag	2008	45.23	-73.34
42	14	CC34	N	ECH	2008	45.18	-73.35
32.1	14	CC34	F	WTD	2008	45.33	-73.04
19.1	1	CC403	F	WTD	2008	45.24	-72.97
154	3	CC3	N	RS	2008	45.18	-73.35
128	14	CC34	N	ECH	2008	45.06	-73.28
qc10.d.467	4	CC4	N	Drag	2010	45.78	-72.08
85	8	CC8	N	ECH	2008	45.12	-73.21
620	1	CC403	L	WFM	2008	45.31	-73.01
61.2	1	CC403	F	WTD	2008	45.17	-73.04
33	3	CC3	L	WFM	2008	45.18	-73.35
264	59	CC7	F	Drag	2008	45.12	-73.21
25.4	4	CC4	M	WTD	2008	45.06	-73.22
102	59	CC7	L	ECH	2008	45.12	-73.21
QC12a-085	3	CC3	M	Drag	2012	45.55	-73.32
QC12a-138	3	CC3	N	Drag	2012	45.55	-73.32
QC12a-154	4	CC4	N	Drag	2012	45.28	-72.98
QC12a-163	12	CC12	N	Drag	2012	45.28	-72.98
QC12a-168	1	CC403	N	Drag	2012	45.28	-72.98
QC12a-172	1	CC403	N	Drag	2012	45.28	-72.98
QC12a-175	3	CC3	N	Drag	2012	45.28	-72.98

QC12a-176	1	CC403	N	Drag	2012	45.28	-72.98
QC12a-192	1	CC403	N	Drag	2012	45.28	-72.98
QC12a-198	12	CC12	N	Drag	2012	45.28	-72.98
QC12a-206	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-207	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-212	315	CC34	N	Drag	2012	45.28	-72.98
QC12a-218	1	CC403	N	Drag	2012	45.28	-72.98
QC12a-222	643	CC12	N	Drag	2012	45.28	-72.98
QC12a-228	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-249	19	CC19	N	Drag	2012	45.28	-72.98
QC12a-251	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-254	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-258	644	ST644	N	Drag	2012	45.28	-72.98
QC12a-266	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-275	1	CC403	N	Drag	2012	45.28	-72.98
QC12a-280	36	CC36	N	Drag	2012	45.28	-72.98
QC12a-285	12	CC12	N	Drag	2012	45.28	-72.98
QC12a-286	16	CC16	N	Drag	2012	45.28	-72.98
QC12a-293	3	CC3	N	Drag	2012	45.28	-72.98
QC12a-294	12	CC12	N	Drag	2012	45.28	-72.98

The genotypes shown here are sequence types (ST) identified using the MLST method and the corresponding clonal complex (CC), identified with the single locus variant criterion (as described in [10]), is also shown. Tick types N, F and M are nymphal, adult female and adult male ticks, respectively. Tick sources are Drag for questing ticks collected by drag sampling, and ECH, RS, WFM, and WTD for feeding ticks from eastern chipmunk, red squirrel, white-footed mouse, white-tailed deer respectively.

Landscape mapping

Landscape of the study region was classified using cloud-free Landsat 8 OLI (Operational Land Imager) images available in the USGS database (http://www.usgs.gov/). Nine images were used to create a landscape map, which was developed using Geomatica 2016 (PCI Geomatics, Markham, Ontario). Three Landsat scenes of the study area, and three images per scene from different seasons (spring, summer, fall) and years (2013, 2014, 2015) were examined to optimise landscape classification accuracy. We assumed that landscape had not changed significantly between the sample collections in 2008-2012 and the landscape evident

in the Landsat scenes. Atmospheric correction was applied to the images and panchromatic (black-and-white) images of 15m resolution were fused with multispectral (colour) images of 30m resolution using the intensity-hue-saturation (IHS) transformation method [28] to obtain a high-resolution color image of 15m resolution. A supervised maximum likelihood classification [29], an expert-driven classification supported by a maximum likelihood algorithm, was used to distinguish five landscape classes: (1) agricultural land, (2) urban areas, (3) open water, (4) forests, (5) open wetland and (6) "other" (unclassifiable regions due to cloud, shadow or snow cover). Vegetation indices and new divergence indices were used in the classification process to improve the separation between the spectral signatures of the different classes [30]. The landscape classes were selected on the basis of relevance for distinguishing the natural forest habitats of the small mammal hosts and *I. scapularis* tick vector of *B. burgdorferi* from other less suitable habitats. The final map is shown in Fig 1.

Network analysis and landscape connectivity

Conefor 2.6 [31, 32] was used to assess landscape connectivity of the landscape map as defined by suitable habitat patches (i.e. forest) and corridors linking them. The pairwise least-cost method was used to define the functional connection between patches according to possible movement distances of small mammals, and the resistance of different landscape types to mammal host movement [33, 34]. Distances matrices for habitat patches and nodes (i.e. the sampling sites) were computed using Graphab 2.0.5 [34] after building the landscape graph using the 8-neighbourhood option which considers that two pixels of habitat linked by a corner belong to the same patch. It was assumed that, at the geographic scale of this study, dispersion was equally possible in any direction. Pixels of the map were parameterized with impedance values that ranged from the most permeable for woodland rodents (i.e. forest with a resistance = 1) to least permeable (i.e. urban areas with a resistance = 10,000) according to values available for the white-footed mouse [21]. The three small mammal species from which tick samples were collected for this study (*P. leucopus, T. striatus* and *T. hudsonicus*) have similar annual dispersal distances of approximately 1000m [35, 36, 37, 38].

Conefor 2.6 assesses the connectivity between nodes by comparing the distance between the nodes and the species-specific threshold dispersal distance for the investigated organism - if

the distance between two nodes is shorter than the assigned threshold dispersal distance, nodes are considered as completely connected [32, 39]. Because we do not have information about the potential dispersal distance of all mammals occurring in the study region, we chose a mean dispersal distance of 938m to calculate connectivity. However, connectivity estimates were also estimated for dispersal distances of 500m and 2000m to explore if this affected the results. The probability of direct dispersal between nodes was calculated as a decreasing exponential function of internode distances (i.e. the mean distance corresponds to a 0.5 of probability). In this study, we employed the most commonly used graph-based measure of connectivity which is the probability of connectivity (PC) [39, 40]. The maximum probability of movement across all potential paths between patches, including directly linked patches and/or patches separated by a mosaic of intermediate sub-patches was used to calculate PC [41].

PC ranges between 0 and 1 and is calculated as:

$$PC = \frac{1}{A_L^2} \sum_{i=1}^{n} \quad \sum_{j=1}^{n} a_i a_j p_{ij}^*$$

where p_{ij}^* is defined as the maximum product probability of all possible paths between nodes i and j. and a_i iand a_j are the areas of the habitat patches i and j. n is the total number of nodes in the landscape and A_L Lis the area of study region. PC is based on probabilistic connection model, where a certain probability of dispersal between the two patches considered (p_{ij}) characterizes the links between nodes i and j in the graph.

The p_{ij} is formulated as follow:

$$p_{ij} = e^{-kd_{ij}}$$

Where k is a constant and dij is the least-cost distance value. k is a parameter controlling the rate of decline in the dispersal probability as distance increases [42]:

$$k = {-\log(\theta)}/{d}$$

Where θ is the mean value of the dispersal probability for a particular mean dispersal distance d.

Statistical analysis

The null hypothesis H_0 was that connectivity of forest habitat does not influence the pattern of occurrence of different strains of *B. burgdorferi* in the area of study. The area of study, at the time the samples were collected, represents the front of invasion of *B. burgdorferi* into southern Canada from the US [43]. For that reason our aim was to explore landscape connectivity in the context of initial spread of *B. burgdorferi*, rather than exploring isolation-by-distance, which would be more appropriate to explore once *B. burgdorferi* populations approach equilibrium. In studies of this type it is common to explore the genetic distance between two samples as a continuous outcome (e.g. [44]). However, the MLST strain typing method used here involves analysis of housekeeping genes that would not be expected to vary over the timescale of sampling so the strain type for each sample was considered a categorical variable. We considered that H_0 could be rejected if the likelihood that two samples, collected from two different field sites, were of the same *B. burgdorferi* ST was significantly greater the more highly connected the sites were. To test this we first built logistic regression models in which the outcome variable was whether or not two samples collected at two different sites were of the same ST. The main explanatory variable explored was the PC between the sites, but we also explored whether simple Euclidean distance between sites was a confounding factor as it could be expected that sites closer together may be more likely to have greater connectivity. In the first model, the outcome variable was whether or not two samples were of the same ST (value = 1 if two samples were the same ST, value = 0 if the two samples were different STs). Then similar logistic regression models were created for each ST in the database (value = 1 if two samples were of the same specific ST, value = 0 if two samples were not of the same specific ST). A total of 10 models were constructed for each ST in the dataset for which there were more than one sample. Again, the PC between the sites and Euclidian distance between sites were investigated as possible explanatory variables for the

likelihood that samples in different sites were of the same ST. For all the data matrices, comparisons amongst samples from the same sites were eliminated because these may be particularly likely to be of the same ST due to processes such as founder events [8]. Linearity of associations between explanatory and outcome variables was explored by Lowess smoothed plots of the relationship between PC and the probability that samples were the same ST. Informed by this, PC was converted to a categorical variable. Generalized linear mixed effect models with a logit link function in R version 3.3.2 [45] were used. The temporal separation of samples (as 8 categories for each possible inter year gap: 2008-2008, 2008-2010, 2008-2009, 2008-2012, 2009-2010, 2009-2012, 2010-2012, 2012-2012) was investigated as a categorical random effect because in a zone of invasion there may be dynamic changes in strain frequencies associated with processes such as founder events that may in some way confound spatial analyses [8]. The level of significance of the final models was $P < 0.05$.

One possible confounding factor is that as connectedness between patches increases, these patches form a 'forest meta-patch' where biodiversity (including that of *B. burgdorferi* STs) would be greater [46]. Because sites were not sampled at the same times with the same intensity, precise comparisons of ST diversity amongst sites cannot be made. However to provide a simple measure of possible effects of connectivity of patches and richness of strains, the connectedness of each sampling site to others was calculated using Conefor 2.6 by summing the probability of direct dispersal (p_{ij}) between one patch i and all other patches directly connected to it [32, 34] and compared with the richness of strains at that site. Only sites sampled by drag sampling and rodent capture were included in this analysis, and four sites were considered as part of the same patch because of their close geographic location and high connectivity. For the comparison, Wilcoxon's rank test was used to assess whether the level of connectivity was significantly different between two groups of sites, one with high (> 2 STs) and one with low (≤ 2 STs) ST richness. The level of significance was $P < 0.05$.

Results

PC was categorised into three groups according to the 33rd and 66th centiles, a "low PC" group with PC values of 3.3 x 10^{-5} to 21.1 x 10^{-5}, a "medium PC" group with PC values of 7.3465 x 10^{-3} to 0.561, and a "high PC" group with PC values > 0.561. These groups were the same when PC was estimated using rodent dispersal distances of 500 and 2000m. The likelihood that two samples were of the same ST was significantly greater the higher was the PC value (Table 2; Fig 2). The likelihood that two samples were of the same ST did not vary significantly with the Euclidean distance between sites (P > 0.1). Categorisation of the data into three groups according to the 33rd and 66th centiles of the PC values obtained using dispersal distances of 500m and 2000m resulted in groups that were identical to those obtained when PC was estimated at 938m.

Table 2: Results of mixed effects logistic regression model to investigate the relationship between the likelihood that samples at different sites are of the same ST and landscape connectivity (PC).

	Odds ratio	**95% CI**	**z**	**P**
Low PC	reference			
Medium PC	3.357	1.121 – 10.054	2.16	<0.05
High PC*	5.183	1.667 – 16.119	2.89	<0.01

The model accounted for year of sampling as a random effect. Values for PC were categorized. CI = 95% confidence interval. * The probability that samples at different sites are of the same ST was significantly higher in the High PC group versus that for the Medium PC group ($x^2 = 8.37$, $P < 0.05$).

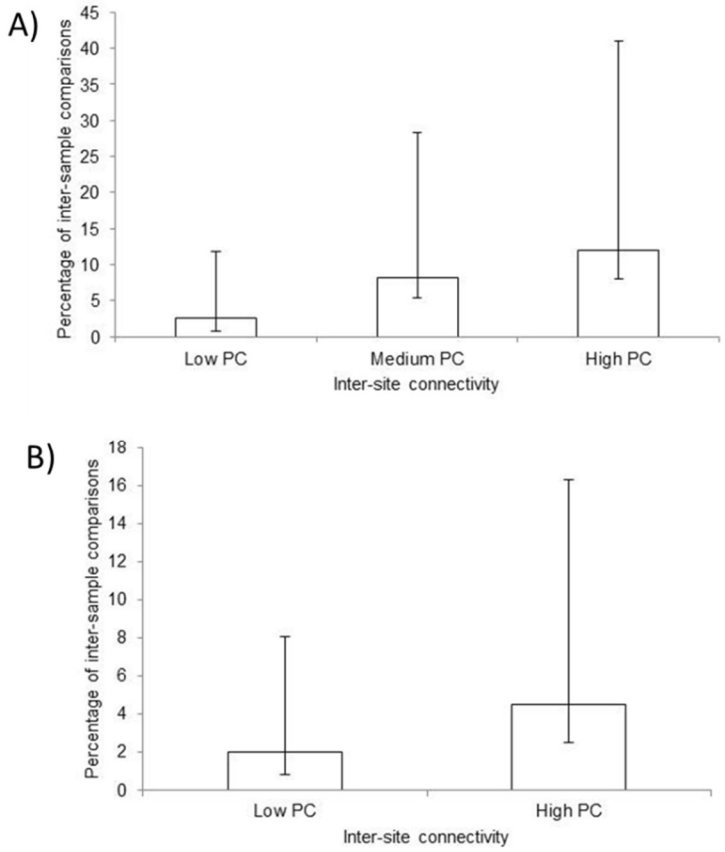


Figure 2: The percentages of sample comparisons that were of the same ST.

In panel A the percentages of sample comparisons that were of the same ST (with exact 95% confidence intervals) when the inter-site connectivity (PC) was low, medium and high as defined in the text is shown. In panel B, the percentage of sample comparisons (with exact 95% confidence intervals) that were ST1 when the inter-site connectivity (PC) was low and high as defined in the text is shown.

In models in which the outcome variable was whether or not two samples were of the same and specific ST, PC was categorized into two groups around the 50th centile (one "low PC" group with PC $\leq$ 0.4422, and one "high PC" group with PC > 0.4422) due to the low number of positive results when exploring the individual STs. The likelihood that two samples were of ST1 was significantly greater with greater PC, and significantly lower the greater the Euclidean distance between the sites (Table 3). However for both ST4 and ST14, the

likelihood that two samples were the same ST was significantly greater the greater was the Euclidean distance between the sites (respectively, odds ratios = 1.02 and 1.07, 95% confidence intervals = 1.00 – 1.05 and 1.03 – 1.12, $P < 0.05$ and < 0.01).

Table 3: Results of mixed effects logistic regression model to investigate the relationship between the likelihood that samples at different sites were ST1 and the degree of connectivity between sites (PC) and Euclidean distance.

	Odds ratio	**95% CI**	**z**	**P**
Connectivity				
Low PC	reference			
High PC	2.537	1.002 – 6.411	1.97	<0.05
Euclidean distance	0.938	0.029 – 0.973	-3.37	<0.01

The model accounted for year of sampling as a random effect. CI = 95% confidence interval.

Sites with high (> 2) ST richness had significantly higher connectivity than sites with low (≤ 2) ST richness ($P < 0.05$; Fig 3).

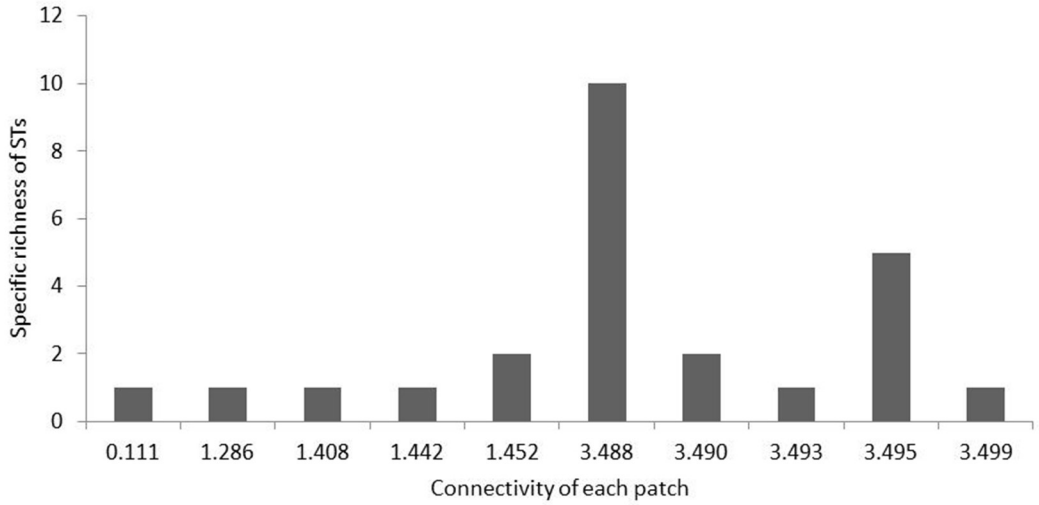


Figure 3: Variation in the richness of STs by estimates of the connectivity for each patch.

Note that sites where ticks were only collected from culled deer were not included in this analysis, and four sites were considered as part of the same patch because of their close geographic location and high connectivity.

Discussion

In this study we found significant association between forest connectivity and the likelihood that two samples were of the same ST. This supports the hypothesis that forest connectivity impacts geographic patterns of *B. burgdorferi* strains dispersal by animal hosts. It is an assumption that the association between connectivity and the patterns of strains is due to effects of connectivity on *B. burgdorferi* host dispersal, but there is much evidence to support this. For example, rodent movements depend on their ability to penetrate habitats [47, 48]; responses to landscape resistance features are similar for the main rodent reservoirs (white-footed mice, Eastern chipmunk and red squirrel: [49, 50, 51, 52]); movement of individuals of the different host species is limited to the same degree by impedance of inter-patch surfaces, and by physical barriers [19, 53, 54, 55]; and forest habitats provide the least impedance to rodent host dispersal in our study region [21].

The relationship between forest connectivity and the likelihood that two samples were of the same ST was not absolutely linear, which is a common finding in spatial patterns associated with landscape connectivity [56, 57], particularly in the context of mammal dispersal. For example the genetic structure of populations of the American Marten (*Martes americana*) has a non-linear relationship with landscape connectivity which has been attributed to dispersal polymorphism due to the landscape connectivity having different effects on different individuals within the species [33]. In our case this non-linearity could in part be explained by dispersal of different STs being affected differently by landscape connectivity, and by other factors determining the frequency of different STs in the study sites.

Samples carrying ST1, which is a member of a clonal complex that has been associated with white-footed mice [10], were more likely to be found at sites that were more connected. Also, samples at sites that were geographically closer together were more likely to be of this strain. This suggests that both higher landscape connectivity and geographic closeness were associated with samples at different sites being more likely to be ST1, which would be consistent with dispersal of this strain by rodents.

Factors other than rates and distances of dispersion may have been operating in determining spatial patterns of occurrence of different strains at the time of sampling. We raised the

possibility that connectivity of forest patches could alter frequencies of different strains by mechanisms other than those affecting dispersal. One possible effect of increased connectivity in a region where *B. burgdorferi* transmission is stable or at equilibrium would be greater diversity (of reservoir hosts and *B. burgdorferi* strains) in connected "meta-patches" than in isolated patches under the premise that greater fragmentation reduces diversity [58]. It is possible that sites belonging to the most highly connected patches did have greater ST richness (Fig. 3), although further prospective studies are needed to confirm this. If ST diversity increases with patch connectedness (and patch connectedness is a measure of belonging, or not, to a metapatch) then the probability that two samples at two sites are the same ST should be lower the more the sites are connected in a metapatch with higher ST richness (Fig 3). We did not observe that, but it could have been a factor that reduced the strength of the relationship we found between connectivity and the likelihood that samples were the same ST.

For two STs (4 and 14), samples at different sites were more likely to be the same ST the more distant were the sites in which they were found. We speculate that this observation may be consistent with immunological factors, in particular negative frequency-dependent selection, affecting the frequency of different strains [58]. While the genes sequenced in the MLST scheme are not immunogenic to our knowledge, there is considerable linkage disequilibrium in the genome of *B. burgdorferi* so it is likely that effects of herd immunity on immunogenic protein allele frequencies are reflected in the frequencies of different MLST STs [59, 60, 61]. Certainly ST4 and ST14 occurred at more of the patches that had low connectivity and high mean Euclidean distance to other sites, than patches that had high connectivity and low mean Euclidean distance to other sites (Fig 4). Patches with higher connectivity had higher ST richness (Fig 3) that may enhance the possibility of hosts acquiring multiple strain infections and cross-reactive antibodies, or as yet undetermined non-immunological inter-strain interactions in the host [62]. This may be particularly so in hosts such as chipmunks that are much longer lived than mice, acquire more infective tick bites in their lifetimes, and with which ST14 has been associated [10]. In contrast the association of ST1 with mice that are relatively short-lived may mean that this strain is less affected by negative frequency-dependent selection associated with herd immunity, resulting in dispersal being a more dominant factor in determining the spatial pattern of this strain.

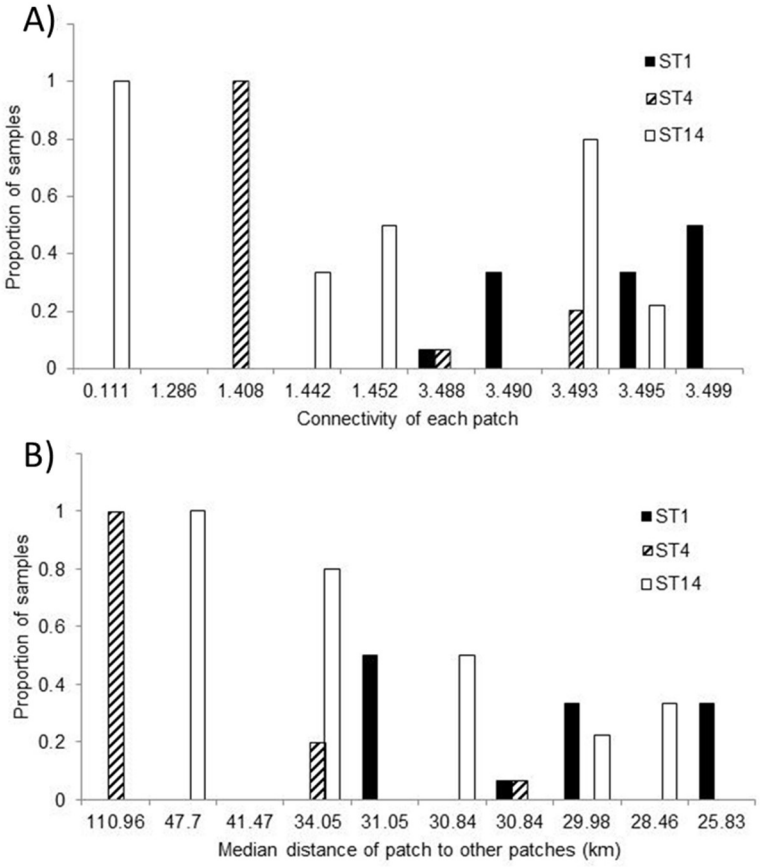


Figure 4: The proportions of samples that were ST1, 4 and 14 by patch connectivity and Euclidean distance.

The proportions of samples that were ST1, 4 and 14 by patch connectivity is shown in panel A. Note that sites where ticks were only collected from culled deer were not included in this analysis, and four sites were considered as part of the same patch because of their close geographic location and high connectivity. The proportions of samples that were ST1, 4 and 14, within the patches are shown in panel B by the median distance of the patch to the other patches.

The generalisability of our findings may be context specific. Our study is in a region of dynamic invasion of *B. burgdorferi* and *I. scapularis* and consequent emergence of Lyme disease in humans, so patterns of strain occurrence associated with dispersal could perhaps be expected to be stronger than in regions where transmission cycles are mature and at

equilibrium. Indeed, differences in gene flow patterns of *I. scapularis* tick populations between the areas studied here, where there is evidence of recolonisation-extinction cycles [63], and more mature endemic populations [64] have been observed. One disadvantage of examining spread patterns of *B. burgdorferi* in a zone and at the time of invasion is that the prevalence of infected tick samples is very low [43] so a limited sample size was available for statistical analysis.

Estimates of genetic distance are useful for exploring isolation by distance. However, this type of analysis is not possible using the data in our study because the time and place of our study is the point of invasion of strains from the US, and the STs in our study are identical to STs present in source populations in the northeastern US [27]. At first sight, other sequences used for strain typing *B. burgdorferi*, such as the *rrs-rrlA* (IGS region and outer surface protein C-encoding sequences (*ospC*), which are generally more variable than each individual gene of the MLST system, may seem more suitable for such an analysis. However, the combination of several loci in the MLST scheme has higher resolution than the IGS or *ospC*. Additionally, in the Northeastern USA these two loci may form strong phylogenetic clusters that are used as categories (IGS types and *ospC* types) in strain typing *B. burgdorferi* (e.g. [9, 65] that correlate with strain typing by MLST due to extensive linkage-disequilibrium in the *B. burgdorferi* genome in this region [9]. Furthermore, for all these strain-typing sequences the frequency of different strains varies geographically in North America, but strains from the same geographic area do not cluster together in the phylogenetic trees (reviewed in [11]). The use of whole genome sequencing in the future may well assist more detailed analysis of potential dispersion patterns by examination of single nucleotide polymorphisms in a large number of orthologous loci.

Our findings are consistent with those in Europe where the distribution of different genotypes of the rodent specialist species *B. afzelii* is highly dependent on landscape (or at least landscape barriers), while the same is not true for the bird specialist *B. garinii* [22]. Our findings also support the hypothesis that population expansions and dispersion of terrestrial hosts associated with past glacial-interglacial climate changes coupled with host association, may have driven the current phylogeographic patterns of *B. burgdorferi* in North America [11, 66]. The range of the reservoir host population does indeed appear to be expanding north at present with a warming climate [7].

Conclusions

Our study suggests that patterns of *B. burgdorferi* strain occurrence are determined at least in part by forest connectivity, and suggests that landscape connectivity, by impacting patterns of host dispersal i) determines in part the phylogeography of *B. burgdorferi* at a number of spatial scales; and ii) is an important factor in any attempt to predict trajectories of spread of the bacterium at a range of scales from local to continental.

Acknowledgments

This study was funded by the Public Health Agency of Canada.

References

1. Gubler DJ, Reiter P, Ebi KL, Yap W, Nasci R, Patz JA. Climate variability and change in the United States: potential impacts on vector- and rodent-borne diseases. Environ Health Perspect. 2001;109: 223-233.
2. Gould EA, Higgs S. Impact of climate change and other factors on emerging arbovirus diseases. Trans R Soc Trop Med Hyg. 2009;103: 109-121.
3. Bueno-Marí R, Jiménez-Peydró R. Global change and human vulnerability to vector-borne diseases. Front Physiol. 2013;4: 158.
4. Bouchard C, Leonard E, Koffi JK., Pelcat Y, Peregrine A, Chilton N, et al. The increasing risk of Lyme disease in Canada. Can Vet J. 2015;56: 693-699.
5. Leighton P, Koffi J, Pelcat Y, Lindsay LR, Ogden NH. Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. J Appl Ecol. 2012;49: 457-464.
6. Ogden NH, Radojevic M, Wu X, Duvvuri VR, Leighton PA Wu J. Estimated effects of projected climate change on the basic reproductive number of the Lyme disease vector *Ixodes scapularis*. Environ Health Perspect. 2014;122: 631-638.

7. Simon JA, Marrotte RR, Desrosiers N, Fiset J, Gaitan J, Gonzalez A, et al. Climate change and habitat fragmentation drive the occurrence of *Borrelia burgdorferi*, the agent of Lyme disease, at the northeastern limit of its distribution. Evol Appl. 2014;7: 750-764.
8. Ogden NH, Mechai S, Margos G. Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. Front Cell Infect Microbiol. 2013;3: 46.
9. Margos G, Gatewood AG, Aanensen DM, Hanincova K, Terekhova D, Vollmer SA, et al. MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. Proc Natl Acad Sci U S A. 2008;105: 8730-8735.
10. Mechai S, Margos G, Feil EJ, Barairo N, Lindsay LR, Michel P, et al. Evidence for Host-Genotype Associations of *Borrelia burgdorferi* Sensu Stricto. PLoS One. 2016;11: e0149345.
11. Ogden NH, Feil EJ, Leighton PA, Lindsay LR, Margos G, Mechai S, et al. Evolutionary aspects of emerging Lyme disease in Canada. Appl Environ Microbiol. 2015;81: 7350-7359.
12. Hanincova K, Mukherjee P, Ogden NH, Margos G, Wormser GP, Reed KD, et al. Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. PLoS One. 2013;8: e73066.
13. Stafford KC 3rd. Oviposition and larval dispersal of Ixodes dammini (Acari: Ixodidae). J Med Entomol. 1992;29:129-32
14. Humair PF. Birds and Borrelia. International Journal of Medical Microbiology. 2002;291: 70-74.
15. Ogden NH, Lindsay LR, Hanincova K, Barker IK, Bigras-Poulin M, Charron DF, et al. Role of migratory birds in introduction and range expansion of *Ixodes scapularis* ticks and of *Borrelia burgdorferi* and *Anaplasma phagocytophilum* in Canada. Appl Environ Microbiol. 2008;74: 1780-1790.
16. Ogden NH, Lindsay LR, Leighton P. Predicting the rate of invasion of the agent of Lyme disease, *Borrelia burgdorferi* in North America. J Appl Ecol. 2013;50: 510-518.

17. Madhav NK, Brownstein JS, Tsao JI, Fish D. A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). J Med Entomol. 2004;41: 842-852.
18. Bennett AF. Habitat corridors and the conservation of small mammals in a fragmented forest environment. Landsc Ecol. 1990;4: 109-122.
19. Bennett AF, Henein K, Merriam G. Corridor use and the elements of corridor quality: Chipmunks and fencerows in a farmland mosaic. Biol Cons. 1994;68: 155-165.
20. Gilbert-Norton L, Wilson R, Stevens JR, Beard KH. A meta-analytic review of corridor effectiveness. Conserv Biol. 2010;24: 660-668.
21. Marrotte RR, Gonzalez A, Millien V. Landscape resistance and habitat combine to provide an optimal model of genetic structure and connectivity at the range margin of a small mammal. Mol Ecol. 2014;23: 3983-3998.
22. Vollmer SA, Feil EJ, Chu CY, Raper SL, Cao WC, Kurtenbach K, et al. Spatial spread and demographic expansion of Lyme borreliosis spirochaetes in Eurasia. Infect Genet Evol. 2013;14: 147-155.
23. Kurtenbach K, De Michelis S, Etti S, Schafer SM, Sewell HS, Brade V, et al. Host association of *Borrelia burgdorferi* sensu lato--the key role of host complement. Trends Microbiol. 2002;10: 74-79.
24. Mitchell MGE, Bennett EM, Gonzalez A, Lechowicz MJ, Rhemtulla JM, Cardille JA, et al. The Montérégie Connection: linking landscapes, biodiversity, and ecosystem services to improve decision making. Ecol Soc. 2015;20: 15.
25. Bouchard C, Beauchamp G, Nguon S, Trudel L, Milord F, Lindsay LR, et al. Associations between *Ixodes scapularis* ticks and small mammal hosts in a newly endemic zone in southeastern Canada: implications for *Borrelia burgdorferi* transmission. Ticks Tick Borne Dis. 2011;2: 183-190.
26. Bouchard C, Leighton PA, Beauchamp G, Nguon S, Trudel L, Milord F, et al. Harvested white-tailed deer as sentinel hosts for early establishing *Ixodes scapularis* populations and risk from vector-borne zoonoses in southeastern Canada. J Med Entomol. 2013; 50: 384-393.
27. Mechai S, Margos G, Feil EJ, Lindsay LR, Ogden NH. Complex population structure of *Borrelia burgdorferi* in southeastern and south central Canada as revealed by phylogeographic analysis. Appl Environ Microbiol. 2015;81: 1309-1318.

28. Nikolakopoulos KG. Comparison of nine fusion techniques for very high resolution data. Photogramm Eng Rem S. 2008;74: 647-659.
29. Jensen JR. Introductory digital image processing: A remote sensing perspective. 3rd edn. Upper Saddle River: Prentice Hall; 2005.
30. Kotchi SO, Brazeau S, Turgeon P, Pelcat Y, Légaré J, Lavigne MP, et al. Evaluation of earth observation systems for estimating environmental determinants of microbial contamination in recreational waters. IEEE J-STARS. 2015;8: 3730-3741.
31. Saura S, Pascual-Hortal L. A new habitat availability index to integrate connectivity in landscape conservation planning: comparison with existing indices and application to a case study. Landsc Urban Plan. 2007;83: 91-103.
32. Saura S, Torné J. Conefor Sensinode 2.2: A software package for quantifying the importance of habitat patches for landscape connectivity. Environ Modell Softw. 2009;24: 135-139.
33. Broquet T, Ray N, Petit E, Fryxell JM, Burel F. Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). Landsc Ecol. 2006;21: 877-889.
34. Foltête JC, Clauzel C, Vuidel G. A software tool dedicated to the modelling of landscape networks. Environ Modell Softw. 2012;38: 316-327.
35. Krohne DT, Dubbs BA, Baccus R. An analysis of dispersal in an unmanipulated population of *Peromyscus leucopus*. Am Midl Nat. 1984;112: 146-156.
36. Loew SS. Sex-biased dispersal in eastern chipmunks, *Tamias striatus*. Evol Ecol. 1999;13: 557-577.
37. Sun C. Dispersal of young in red squirrels (*Tamiasciurus hudsonicus*). Am Midl Nat. 1997;138: 252-259.
38. Haughland DL, Larsen KW. Exploration correlates with settlement: red squirrel dispersal in contrasting habitats. J Anim Ecol. 2004;73: 1024-1034.
39. Saura S. Measuring connectivity in habitat mosaics: the equivalence of two existing network indices and progress beyond them. Community Ecol. 2010;11: 217-222.
40. Ricotta C, Stanisci A, Avena GC, Blasi C. Quantifying the network connectivity of landscape mosaics: a graph-theoretical approach. Community Ecol. 2000;1: 89-94.

41. Saura S, Vogt P, Velázquez J, Hernando A, Tejera R. Key structural forest connectors can be identified by combining landscape spatial pattern and network analyses. Forest Ecol Manag. 2011;262: 150-160.
42. Clauzel C, Foltête JC, Girardet X, Vuidel G. Graphab 2.0 user manual; 2016.
43. Ogden NH, Bouchard C, Kurtenbach K, Margos G, Lindsay LR, Trudel L, et al. Active and passive surveillance and phylogenetic analysis of *Borrelia burgdorferi* elucidate the process of Lyme disease risk emergence in Canada. Environ Health Perspect. 2010;118: 909-914.
44. Davis MA, Hancock DD, Besser TE, Rice DH, Hovde CJ, Digiacomo R, et al. Correlation between geographic distance and genetic similarity in an international collection of bovine faecal *Escherichia coli* O157:H7 isolates. Epidemiol Infect. 2003;131: 923-930.
45. R Core Team. A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. 2015. Available at http.www.R-project.org
46. Jones NT, Germain RM, Grainger TN, Hall AM, Baldwin L, Gilbert B. Dispersal mode mediates the effect of patch size and patch connectivity on metacommunity diversity. J Ecol. 2015;103: 935-944.
47. McRae BH, Nürnberger B. Isolation by resistance. Evolution. 2006;60: 1551-1561.
48. Zeller KA, McGarigal K, Whiteley AR. Estimating landscape resistance to movement: a review. Landsc Ecol. 2012;27: 777-797.
49. Kozakiewicz M. Habitat isolation and ecological barriers-the effect on small mammal populations and communities. Acta Theriol. 1993;38: 1-30.
50. Nupp TE, Swihart RK. Landscape-level correlates of small-mammal assemblages in forest fragments of farmland. J Mammal. 2000;81: 512-526.
51. Bakker VJ. Microhabitat Features Influence the Movements of Red Squirrels (*Tamiasciurus hudsonicus*) on Unfamiliar Ground. J Mammal. 2006; 87: 124.
52. Marzluff J, Shulenberger E, Endlicher W, Alberti M, Bradley G, Ryan C, et al. Urban ecology: An international perspective on the interaction between humans and nature. New York: Springer US; 2008.

53. Merriam G, Kozakiewicz M, Tsuchiya E, Hawley K. Barriers as boundaries for metapopulations and demes of *Peromyscus leucopus* in farm landscapes. Landsc Ecol. 1989;2: 227-235.
54. Merriam G, Lanoue A. Corridor use by small mammals: field measurement for three experimental types of *Peromyscus leucopus*. Landsc Ecol. 1990;4: 123-131.
55. Bowne DR, Bowers MA. Interpatch movements in spatially structured populations: a literature review. Landsc Ecol. 2004;19: 1-20.
56. Rudnick D, Ryan SJ, Beier P, Cushman SA, Dieffenbach F, Epps C, et al. The role of landscape connectivity in planning and implementing conservation and restoration priorities. Issues Ecol. 2012;16: 20.
57. Grêt-Regamey A, Rabe SE, Crespo R, Lautenbach S, Ryffel A, Schlup B. On the importance of non-linear relationships between landscape patterns and the sustainable provision of ecosystem services. Landsc Ecol. 2014;29: 201-212.
58. Allan BF, Keesing F, Ostfeld RS. Effect of forest fragmentation on Lyme disease risk. Conserv Biol. 2003;17: 267–272.
59. Qiu WG, Bosler EM, Campbell JR, Ugine GD, Wang IN, Luft BJ, et al. A population genetic study of *Borrelia burgdorferi* sensu stricto from eastern Long Island, New York, suggested frequency-dependent selection, gene flow and host adaptation. Hereditas 1997;127:203-16.
60. Dykhuizen D, Brisson D. Evolutionary genetics of *Borrelia burgdorferi* sensu lato. In: D.S. Samuels, J.D. Radolf and J.D. Caister editors, Borrelia--Molecular biology, host interaction and pathogenesis. Norfolk: Academic Press; 2010. pp27-54.
61. Kurtenbach K, Hanincová K, Tsao JI, Margos G, Fish D, Ogden NH. Fundamental processes in the evolutionary ecology of Lyme borreliosis. Nat Rev Microbiol. 2006;4: 660-669.
62. Devevey G, Dang T, Graves CJ, Murray S, Brisson D. First arrived takes all: inhibitory priority effects dominate competition between co-infecting *Borrelia burgdorferi* strains. BMC Microbiol. 2015;15: 61.
63. Leo SS, Gonzalez A, Millien V. The genetic signature of range expansion in a disease vector-the black-legged tick. J Hered. 2016;esw073.

64. Khatchikian CE, Prusinski MA, Stone M, Backenson PB, Wang IN, Foley E, et al. Recent and rapid population growth and range expansion of the Lyme disease tick vector, *Ixodes scapularis*, in North America. Evolution. 2015;69: 1678-1689.
65. Bunikis J, Garpmo U, Tsao J, Berglund J, Fish D, Barbour AG. Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. Microbiology. 2004;150:1741-1755.
66. Margos G, Tsao JI, Castillo-Ramírez S, Girard YA, Hamer SA, Hoen AG, et al. Two boundaries separate *Borrelia burgdorferi* populations in North America. Appl Environ Microbiol. 2012; 78: 6059-6067.

VI. Discussion générale

1. Résumé du problème et de la méthode d'étude

Notre étude avait pour but i) de mieux comprendre l'épidémiologie moléculaire de *B. burgdorferi* au Canada; ii) d'identifier sa relation avec les hôtes réservoirs; et iii) de déterminer le rôle de ces derniers dans la dispersion des souches de *B. burgdorferi* ?

Dans cette étude, nous avons exploité une partie du génome de *B. burgdorferi* pour identifier des variables écologiques nécessaires pour comprendre l'écologie contemporaine de la maladie de Lyme au Canada. En effet, l'une des conclusions des études de Kurtenbach et al. (2006) et d'Ogden et al. (2013b), portant sur les processus fondamentaux qui définissent l'écologie évolutive de la maladie de Lyme, est que des signatures d'évènements anciens peuvent être détenues dans le génome de ce spirochète. Ces empreintes génétiques sont exploitables par des méthodes statistiques (Kurtenbach et al. 2006).

Pour identifier et caractériser les souches de *B. burgdorferi*, nous avons utilisé trois approches moléculaires complémentaires : MLST, ospC et IGS. Ce choix est motivé par le fait que le génome de *B. burgdorferi* est assez riche, composé d'un génome de base (chromosome linéaire) et d'un génome accessoire (plasmides). L'utilisation des gènes de ménage avec MLST a permis d'exploiter la variation génétique au niveau de la partie du génome la plus stable (c'est la méthode la plus recommandée jusqu'à présent pour typer *B. burgdorferi*) (Margos et al., 2011). L'utilisation des deux autres approches a permis d'exploiter d'autres parties du génome plus affectées par la sélection naturelle; ce sont les techniques les plus utilisées pour des études de pathogénicité d'isolats de *B. burgdorferi* isolés chez l'humain avant l'arrivée de la méthode MLST (Qiu et al., 2004; Hanincova et al., 2008a). Ceci a permis d'investiguer des associations entre ces loci, car des études ont rapporté que les allèles de plusieurs souches de *B. burgdorferi* seraient en déséquilibre de liaison (Brisson et al., 2012).

Les échantillons d'ADN de *B. burgdorferi* ont été extraits à partir des tiques *I. scapularis*. Ces tiques proviennent de plusieurs sources. Elles ont été collectées soit dans l'environnement, soit prélevées sur des rongeurs piégés et/ou sur des cerfs de virginie chassés, soit à partir de biopsies de cœur des rongeurs. Les prélèvements et les collectes ont été réalisés par les équipes de l'ASPC lors des campagnes de surveillance active (2001 à 2013). Étant

donné que la distinction morphologique des souris sylvestres et des souris à pattes-blanches est difficile, le laboratoire National de microbiologie a utilisé la PCR pour confirmer l'identification des souris.

Notre choix des données de surveillance active dans les analyses statistiques est fondamental, car elles apportent la précision de géo-référencement des données génétiques, nécessaire pour investiguer et modéliser efficacement les variations génétiques intra-spécifiques à l'échelle spatiale (Thomassen et al., 2010). C'est l'un des aspects principaux qui distingue cette étude des précédentes qui se sont intéressées à la diversité génétique de *B. burgdorferi* au Canada (Ogden et al., 2011).

Pour caractériser le paysage et identifier le milieu naturel du vecteur et des hôtes réservoirs de *B. burgdorferi*, nous avons exploité des images de télédétection spatiale. Différentes sources de données d'imagerie existent telles que celles produites par le satellite «QuickBird» que Marrotte et al. (2014) ont utilisé. Cependant, pour pouvoir détecter et prédire efficacement les corridors, nous avons utilisé les images Landsat 8 qui sont beaucoup plus précises (Taylor et al., 2011). La disponibilité des images de haute qualité, d'une façon périodique et renouvelée, permettra de développer des modèles prédictifs plus précis, cette précision étant nécessaire pour optimiser les calculs de la connectivité du paysage (Boyle et al., 2014). Dans le futur, l'utilisation des images avec une résolution < 5 m, telles que celles produites par le satellite IKONOS permettra d'améliorer la détection de fins changements dans les habitats naturels, nécessaires pour mesurer les effets de la connectivité du paysage sur les mouvements des hôtes de *B. burgdorferi* (Boyle et al., 2014).

Beaucoup d'efforts ont été consentis pour exploiter d'une façon optimale les données. Les séquences d'ADN de *B. burgdorferi* ont été analysées avec de multiples méthodes complémentaires qui sont adaptées aux données MLSTs (arbre phylogénétique bayésien, BAPS et goeBURST). Chaque méthode est utilisée pour un objectif précis, tout en donnant des résultats cohérents avec les autres méthodes (p.ex. les groupes de BAPS sont cohérent avec les clades de l'arbre phylogénétique et les CCs de goeBURST). Les analyses statistiques ont été menées avec des méthodes multivariées (p.ex. analyse de correspondance) et des méthodes multi-variables (p.ex. modèles de régression logistiques mixtes avec effet aléatoire) qui ont permis d'obtenir des résultats consistants. Cette méthodologie a été validée par des experts internationaux qui sont des coauteurs dans les différentes publications.

2. Examen et discussion des principaux résultats de l'étude

Pour comprendre l'épidémiologie actuelle de la maladie de Lyme au Canada, nous avons tenté de retracer l'histoire évolutive de *B. burgdorferi* en exploitant la variation génétique que montrent son génome de base (chromosome linéaire). Même si le nombre de gène de ménage que possède *B. burgdorferi* est beaucoup moins important que le nombre de gènes de plasmides (Qiu et al., 2004), les 8 gènes de ménage choisis dans le schéma MLST sont suffisamment informatifs et discriminatoires (Margos et al., 2011).

2.1. Histoire évolutive et diversité de *B. burgdorferi* au Canada

Notre étude de Mechai et al. (2015) est en accord sur plusieurs points avec d'autres études récentes qui se sont basées également sur les mêmes gènes de ménage de MLST pour investiguer l'évolution de *B. burgdorferi* en Amérique du Nord (Margos et al., 2008; Hoen et al., 2009; Ogden et al., 2011; Margos et al., 2012; Hanincova et al., 2013; Jungnick et al., 2015).

Le premier point est que l'analyse de l'occurrence des souches communes produit une image phylogéographique miroitée au Canada par rapport aux Nord-est et au Midwest des ÉU. Cette image suit les principales voies migratoires des oiseaux en Amérique du Nord, c'est-à-dire que les souches présentes dans le Haut-Midwest se retrouvent dans le Sud de Manitoba et le Sud-ouest d'Ontario (voie de migration du Mississippi), et celles qui se trouvent dans le Nord-est des ÉU se retrouvent dans le Sud-est Canadien allant du Sud-est de l'Ontario, au Sud du Québec et jusqu'au Sud des Provinces Maritimes (voie de migration de l'Atlantique) (Morshed et al., 2005; Ogden et al., 2013a). Ceci confirme le rôle majeur des oiseaux migrateurs qui introduisent ces souches vers le Sud Canadien.

Le second point est que de façon surprenante, nous avons constaté que dans l'ensemble, seulement environ un cinquième des STs étaient communs aux deux pays et la moitié des STs ne se produisent qu'aux ÉU. Ceci peut être l'œuvre de patrons locaux spécifiques à chaque région tels que le climat et l'utilisation des terres (Humphrey et al., 2010), et à la radiation adaptative de *B. burgdorferi* aux ÉU (Hanincova et al., 2008b).

Un autre point d'accord avec les études d'Ogden et al. (2011) et de Margos et al. (2012) est que le patron géographique semble être complexe, car plusieurs clades formés sont constitués de souches qui viennent de plusieurs zones géographiques parfois très éloignées, et

qui ne correspondent pas aux corridors aériens des oiseaux migrateurs (p.ex. ST17 de la Californie qui forme un clade avec la ST536 des Maritimes : figure 2 dans Mechai et al., 2015). Ceci indique que l'évolution de *B. burgdorferi* sur l'arbre phylogénétique n'est pas basée clairement sur une isolation géographique en Amérique du Nord.

Margos et al. (2012) avaient détecté trois structures génétiques de *B. burgdorferi* qui sont réparties respectivement dans trois zones géographiques, le Nord-est, le Midwest et l'Ouest des ÉU. Cette structuration a été révélée en utilisant une méthode statistique dite «Wombling» qui exploite les fréquences alléliques localement et leurs coordonnées géographiques pour identifier les limites géographiques des variations génétiques (Crida et Manel, 2007). Cependant, étant donné que ces trois structures ne sont pas clairement identifiables sur l'arbre phylogénétique, ces auteurs ont attribué cette structure génétique à des processus locaux qui peuvent être affectés par des barrières géographiques (p.ex. les grandes plaines agricoles entre le Nord-est et le Midwest).

De même, la diversité élevée de *B. burgdorferi* que nous avons noté dans le Sud de Manitoba et le Nord-ouest d'Ontario est conforme aux études antérieures aux ÉU qui suggèrent que la richesse de *B. burgdorferi* est plus élevée dans le Haut-Midwest (c'est-à-dire immédiatement au Sud du Manitoba) que dans le Nord-est (c'est-à-dire immédiatement au Sud-est de l'Ontario, du Québec et des Maritimes) (Hanincova et al., 2013).

Notre étude confirme également que la variation génétique qui se produit au niveau des gènes de ménage est beaucoup plus due à des mutations qu'à des évènements de transfert horizontal (Margos et al., 2012). Ceci est en accord avec l'hypothèse qui soutient que l'évolution des souches de *B. burgdorferi* se fait très lentement due à l'accumulation des mutations génétiques. En d'autres terme, l'évolution et la diversification de *B. burgdorferi* ss en Amérique du Nord s'est faite sur des milliers, voire des millions d'années (Margos et al., 2008; Hoen et al., 2009). En effet, *B. burgdorferi* est l'une des rares bactéries possédant un génome très stable (Mongodin et al., 2013). La diversification de ce génome s'opère et varie en fonction des processus d'expansion et de contraction successives des populations de tiques et de l'agent pathogène, qui se sont produits au cours de leur histoire respective (Mongodin et al., 2013; Ogden et al., 2013a). Ces événements sont souvent engendrés par des grands changements tels que les changements climatiques anciens (p.ex. périodes glaciaires et interglaciaires) et/ou actuels (p.ex. réchauffement climatique). Ces changements ont un impact

direct et/ou indirect sur l'habitat et la raréfaction/disponibilité des hôtes (Medlock et al., 2013; Ogden et al., 2013b). Aussi, les changements anthropiques jouent un rôle important notamment dans la fragmentation du paysage. À titre d'exemple, l'expansion actuelle d'*I. scapularis* et de *B. burgdorferi* en Amérique du Nord a pu être favorisée par des perturbations majeures qu'a subit le paysage telles que la reforestation (à la suite de l'abandon des terres agricoles à l'époque moderne) (Margos et al., 2011) après des décennies de déforestation (à l'arrivée des Européens à l'époque Post-Colombien : Wood et Lafferty 2013).

Notre étude démontre que la diversité de *B. burgdorferi* est plus large que ce qu'on pensait auparavant. Elle s'élève à 138 souches, dont 30 sont nouvelles au Canada. Ceci est plutôt attendu quand on augmente la taille d'échantillonnage et qu'on prospecte de nouvelles zones géographiques, comme révélé par l'analyse des courbes de raréfaction de la diversité de *B. burgdorferi* dans Mechai et al. (2016). Ogden et al. (2011) ont aussi rapporté la présence de nouvelles souches au Canada en exploitant des données de surveillance passive. Cependant, la caractéristique des nouvelles souches identifiées dans notre étude montrent des clusters génétiquement très distincts avec respectivement une certaine diversité génétique très localisée géographiquement et qui opèrent dans trois zones géographiques très éloignée (p.ex. Buffalo Point au Manitoba, Long Point en Ontario, Lunenburg à la Nouvelle-Écosse). Il s'agit probablement d'un signal de la présence de populations refuges au Nord des ÉU pendant les périodes glaciaires et interglaciaires, mais qui ne sont pas explorées jusqu'à présent aux ÉU (Provan et Bennett, 2008; Ogden et al., 2015b). En effet, à l'exception de quelques petites populations de souches canadiennes qui sont restées des « singletons », provenant probablement des populations refuges, l'ancêtre prédit pour la plupart des complexes clonaux (CCs) se trouve aux ÉU (Mechai et al., 2015). Par ailleurs, la caractéristique des souches du Long Point en Ontario identifiées dans notre étude, qui présentent des liens avec des souches du MW et du NE des ÉU, montre que cette zone a probablement abrité une population refuge (Ogden et al., 2015a).

L'hypothèse avancée pour expliquer une telle structure est qu'elle pouvait être l'œuvre d'un patron écologique, en dépit du fait que *B. burgdorferi* est connue pour être généraliste (Kurtenbach et al., 2002; Kurtenbach et al., 2006; Vollmer et al., 2011; Ogden et al., 2013a; Ogden et al., 2015a). La radiation adaptative de *B. burgdorferi* en Amérique du Nord pourrait l'amener à se spécialiser pour infecter certaines espèces d'hôtes dans les périodes d'expansion

des populations (Kempf et al., 2009; Ogden et al., 2013a). Dans une telle situation, le développement d'une association souche-hôte est en général considéré comme un avantage pour les pathogènes comme *B. burgdorferi* (Kurtenbach et al. 2006). Cette association peut limiter le flux génétique entre les populations de *B. burgdorferi* isolées par leurs hôtes (Margos et al., 2012). Ceci peut expliquer en partie la structure des clades et des CCs présentés dans notre étude de Mechai et al. (2015). De même, cela pourrait aussi expliquer le faible flux génétique rapporté auparavant entre les populations de *B. burgdorferi* du Nord-est et du Midwest, les grandes plaines agricoles pouvant jouer un rôle de barrière aux mouvements des hôtes (Margos et al., 2012).

En conclusions, ces résultats sont cohérents avec l'hypothèse selon laquelle les clusters phylogénétiques ne sont pas associés géographiquement, pouvant être engendrés par des processus écologiques tels que l'association des souches aux hôtes. Par conséquent, l'image phylogéographique peut être le résultat i) des expansions passées des populations de *B. burgdorferi* lorsque les conditions environnementales étaient propices à l'adaptation radiative et aux associations des souches aux hôtes; ii) à la dispersion de ces hôtes quand l'environnement leur permettait d'étendre leurs populations; iii) des contractions de ces populations vers des zones refuges (MW, NE) après que les conditions environnementales soient devenues défavorables pour les tiques, les hôtes, et pour *B. burgdorferi* (p.ex. les périodes glaciaires, les pertes d'habitat).

2.2. Diversité et écologie de *B. burgdorferi*

Comme rapporté précédemment, à notre connaissance, *B. burgdorferi* reste un pathogène généraliste qui peut survivre et être transmis par un large éventail d'espèces de vertébrés (Kurtenbach et al., 2006), tout en restant plus fréquent chez certains hôtes.

2.2.1. Associations hôtes-génotypes

Si latitudinalement, la diversité de *B. burgdorferi* en Amérique du Nord semble être guidée par les mouvements des oiseaux migrateurs et la présence de populations refuges au Nord des ÉU, longitudinalement la fluctuation de ses populations semble être facilitée par les dynamiques locales du vecteur et des hôtes.

Dans ce contexte, nos résultats dans Mechai et al. (2016) montrent des différences significatives dans la richesse génotypique de *B. burgdorferi* parmi les espèces d'hôtes. En effet, l'importance de la diversité de *B. burgdorferi* dans le MW (incluant le Sud de Manitoba et le Nord-ouest d'Ontario) par rapport à sa diversité dans le NE (incluant le Sud-est d'Ontario, le Sud de Québec, et les Maritime) peut être expliquée par la durée du chevauchement de l'activité saisonnière des nymphes infectées et des larves de la tique *I. scapularis*. Cette durée, plus importante dans le Midwest par rapport au Nord-est des ÉU (Gatewood et al., 2009), offre ainsi suffisamment de temps aux tiques pour transmettre davantage de génotypes de *B. burgdorferi* (Gatewood et al., 2009).

L'un des paramètres les plus importants qui détermine le comportement épidémiologique des pathogènes est la durée de l'infectiosité (Gatewood et al., 2009). Cette durée est directement liée à la durée de vie de l'hôte et au temps avant que l'organisme de l'hôte puisse éliminer l'agent pathogène (Gatewood et al., 2009). Ces différences dans la dynamique de l'infection chez l'hôte et le vecteur peuvent avoir des effets particulièrement profonds sur la diversité de *B. burgdorferi* (Hanincova et al., 2008b). De plus en plus d'études suggèrent l'importance de la persistance de l'infection de *B. burgdorferi* dans des hôtes réservoirs pour assurer son maintien à long terme dans l'environnement (Kurtenbach et al., 2006; Ogden et al., 2007; Gatewood et al., 2009). D'autres études ont soulevé la possibilité que la persistance de *B. burgdorferi* tout au long de la vie de certains hôtes puisse ne pas être universelle (Lindsay et al., 1997; Derdakova et al., 2004; Kurtenbach et al., 2006). Ceci révèle la pression qu'exerce le système immunitaire de l'hôte sur *B. burgdorferi*. Durant ce processus, *B. burgdorferi* peut s'adapter pour assurer sa survie (Kurtenbach et al., 2006). Ainsi, durant leur adaptation, certains génotypes pourraient mieux survivre que d'autres dépendamment de l'espèce d'hôte (Hanincova et al., 2008b).

Le fait qu'elle soit plus élevée chez les tamias (Mechai et al., 2016) pourrait être dû à la durée de vie relativement longue de cette espèce. Sa durée de vie peut aller jusqu'à 3 ans (Tryon & Snyder, 1973), alors que les autres espèces de rongeurs dominantes (p.ex. souris à pattes-blanches) n'arrivent que rarement à survivre pendant un an (Schug et al., 1991). Par conséquent, les tamias sont plus exposés aux tiques nymphes que les souris, augmentant ainsi leur chance d'être infectés par une plus grande diversité de génotypes de *B. burgdorferi* (Hanincova et al., 2013).

Pour la première fois, notre étude a clairement identifié des associations entre génotypes de *B. burgdorferi* et certaines espèces d'hôtes en Amérique du Nord (Mechai et al., 2016). Globalement, la souche ST1 (CC403) porte souvent les séquences ospC A et RST1. Les deux sont associées avec la souris à pattes-blanche. Par contre, la souche ST14 (CC34) porte souvent la séquence ospC G qui est associée avec le tamia. Les séquences ospC H et IGS 2D de la souche ST4 sont quant à eux associées avec la souris sylvestre.

Notre étude soutient l'ajustement écologique (en anglais : ecological fitting) de *B. burgdorferi,* qui serait à l'origine des fluctuations des fréquences de ses souches, ce qui reste compatible avec les hypothèses en matière d'adaptation radiative lorsque les conditions (habitat et climat) sont favorables (McCoy et al., 2013). Cependant, ces associations d'hôte-souche peuvent être dues à des caractéristiques des génotypes de nature plus subtile qui augmentent la probabilité de les retrouver ou de les transmettre à partir d'une espèce d'hôte particulière (p.ex. l'adaptation/échappement à l'immunité de l'hôte, infections longues, périodes de transmission longues) (Hanincova et al., 2008b). Ainsi, à long terme, cet ajustement peut amener *B. burgdorferi* à évoluer vers la spécialisation comme en Europe. Cet ajustement a une importante capitale dans l'évolution écologique de *B. burgdorferi* en Amérique du Nord, car il permet d'expliquer la phylogéographie contemporaine de ce pathogène. Les processus écologiques semblent définir mieux les structures des clades, des CCs et des groupes de BAPs que les changements géographiques. En d'autres termes, le patron géographique des souches est consistant avec la dispersion des hôtes qui est limitée par l'habitat, le climat et leur survie dans des zones refuges quand ces conditions sont défavorables.

Ces observations sur les associations d'hôte-souche étaient cohérentes dans toutes les analyses statistiques effectuées (multivariées et multi-variables) et pour les différentes méthodes de génotypage utilisées. De même, elles sont consistantes avec l'étude de la dispersion des souches par les hôtes (chapitre III).

2.2.2. Conséquences de l'adaptation de *B. burgdorferi* pour certains hôtes

Nous avons voulu chercher l'empreinte de ces associations d'hôte-souche observées en remontant l'histoire évolutive de *B. burgdorferi* grâce à la phylogénie. Pour cela, seuls les gènes de ménage sont considérés. L'association avec le tamia rayé est strictement due à des

STs anciens du CC34 (Mechai et al., 2016). En effet, ils sont proches du centre de l'arbre, alors que ceux associés avec les souris sont bien loin du centre de l'arbre montrant des branches plus longues. Ceci indique qu'ils ont évolué plus récemment par rapport à ceux associés aux tamias (Figure 8 dans Mechai et al., 2016).

Des études se basant sur des données fossiles du tamia en Amérique du Nord corroborent d'une façon intéressante cette hypothèse. En effet, elles révèlent que cette espèce était l'une des rares espèces de petits mammifères ayant survécu grâce aux zones refuges nordiques pendant les périodes glaciaires et interglaciaires (Rowe et al., 2004; Rowe et al., 2006). Le tamia a ensuite entrepris une migration Nord-sud suite au retrait de la couche glaciaire (Rowe et al., 2004; Rowe et al., 2006). Les souris par contre, qui sont beaucoup plus sensibles au climat (Simon et al., 2014), ont entrepris plutôt une migration Sud-nord, suite aux conditions climatiques devenues plus favorables vers le Pléistocène tardif (Waters, 1963). Ces évènements laissent croire que, *B. burgdorferi* a survécu également dans les zones refuges nordiques pendant les périodes glaciaires et interglaciaires grâce aux tamias rayés. Des adaptations aux souris se sont faites par la suite lorsque ces espèces de rongeurs ont partagé une histoire commune avec les tamias. Les adaptations aux hôtes peuvent aussi avoir été le moteur des différences de pathogénicité des génotypes de *B. burgdorferi* en Amérique du Nord.

2.3. Rôle des hôtes terrestres dans la dispersion de *B. burgdorferi*

Dans cette partie de l'étude, nous avons exploré grâce à la modélisation géo-spatiale le rôle des mouvements des hôtes terrestres dans la dispersion des différentes souches de *B. burgdorferi* à une échelle locale dans le Sud Canadien. Nous nous sommes basés sur l'hypothèse que, s'il y a une préférence de *B. burgdorferi* pour certaines espèces d'hôtes, on devrait retrouver les mêmes souches en suivant les corridors de mouvement des hôtes.

Nous avons choisi la Montérégie comme zone pilote, car elle est l'une des régions montrant le risque le plus important de contracter la maladie de Lyme au Canada (Rapport de surveillance de la maladie de Lyme et des autres maladies transmises par la tique *Ixodes scapularis* au Québec entre 2004 et 2013, INSPQ). La base de données utilisée dans les analyses statistiques comprend les échantillons de *B. burgdorferi* qui proviennent des sites de la Montérégie où deux méthodes d'échantillonnage (flanelle et piégeage de rongeurs) sont

développées. Les données de tiques collectées sur la végétation sont exploitées également, car la tique en dehors de la saison hivernale est toujours attachée à un hôte ou à la recherche d'un hôte (Gatewood et al., 2009). Sachant qu'elle ne se déplace par elle-même que sur quelques mètres durant toute sa vie (Vail & Smith, 2002), ses déplacement sont quasiment assurés par des hôtes. Dans des études semblables, des tiques à la recherche d'hôtes ont été utilisées pour investiguer des associations de génotypes de *B. burgdorferi* sl avec des espèces d'hôtes (p.ex. comme dans Jacquot et al., 2016). Cependant, dans notre étude, ce groupe de tique a été considéré comme une catégorie et la comparaison avec les autres catégories (espèces d'hôtes) a permis de tester la distribution des fréquences des génotypes entre les tiques (attachées à un hôte et celles à la recherche d'un hôte) (c'est-à-dire si la distribution des génotypes entre toutes ces tiques est aléatoire, alors on s'attend est-ce qu'il n'y a pas d'association avec les hôtes).

Nous avons procédé par des probabilités conditionnelles pour modéliser la chance de retrouver une même souche de *B. burgdorferi* (une tique se fait déplacer par une espèce de rongeur d'un patch «A» à un patch «B», sachant que d'autres individus de cette même espèce de rongeur peuvent transporter d'autres tiques infectées par d'autres souches).

Cette étude a montré que ST1 est significativement associé avec les patches les plus connectés. Cet ST, membre du CC403, avait montré précédemment une association avec la souris à pattes-blanches (Mechai et al., 2016). Cela confirme l'hypothèse selon laquelle la connectivité forestière affecte les modèles géographiques de dispersion des souches de *B. burgdorferi*. Par conséquent, l'association entre la connectivité et les modèles de dispersion des souches peut être due aux effets de la connectivité sur la trajectoire (c'est-à-dire la destination mais non pas les directions cardinales) des hôtes de *B. burgdorferi*. En effet, il est connu que les comportements des rongeurs sont guidés par leur capacité à pénétrer les habitats (Brad et al, 2006; Zeller et al., 2012) d'une façon sécuritaire et à moindre coût en terme d'énergie (p.ex. la souris à pattes-blanches) (Marrotte et al., 2014). Par contre, ST4 et ST14 étaient significativement plus associés avec des patches moyennement éloignés les uns des autres (c'est-à-dire distance Euclidienne). Si on avait moins de chance de les retrouver dans les sites les plus connectés (alors qu'ils sont dans des sites voisins), c'est probablement parce que quelque chose réduit leur présence dans ces sites. Ceci peut être l'effet de la sélection négative qu'exerce le système immunitaire de l'hôte sur la fréquence des souches. Les patches

les plus connectés sont les patches qui renferment une grande diversité génétique de *B. burgdorferi* (figure 3 dans chapitre III). Par conséquent, les hôtes qui vivent dans ces patches auront plus de probabilité d'acquérir une mémoire immunitaire contre plus de souches de *B. burgdorferi* (Devevey et al., 2015). Ce dernier résultat suppose que ST4 et ST14 sont plus sensibles que ST1 aux pressions du système immunitaire de leurs hôtes respectifs. En effet, ces deux STs (ST4 et ST14) font partie des CC4 et CC34 qui ont montré des associations avec respectivement la souris sylvestre et le tamia (Mechai et al., 2016). Ceci est d'autant plus vrai pour ST14, car la durée de vie du tamia est plus longue (Tryon & Snyder, 1973).

Nos résultats concordent avec ceux rapportés en Europe où la répartition des différents génotypes de l'espèce *B. afzelii* est très dépendante du paysage (*B. afzelli* est associée avec les rongeurs), alors que ce n'est pas le cas pour *B. garinii* qui est connue pour être associée avec les oiseaux (Vollmer et al., 2013).

Notre étude ici suggère donc que les modèles de l'occurrence des souches de *B. burgdorferi* au Sud Canadien sont déterminés au moins en partie par la connectivité des forêts. Ce paramètre est donc important pour prédire les trajectoires de cette bactérie et sa gamme de souches sur une échelle locale et probablement sur des échelles plus vastes.

En conclusion, la phylogéographie contemporaine de *B. burgdorferi* en Amérique du Nord est le résultat des variations génétiques de ce pathogène qui occurrent latitudinalement et longitudinalement. Ces variations sont causées par une succession de changements (habitat et climat) que subit *B. burgdorferi* dans son milieu naturel (vecteur et hôte). Étant donné que son milieu naturel est dynamique et sa longue histoire évolutive, *B. burgdorferi* peut garder son aptitude d'infecter une large gamme de vertébrés tout en ayant des préférences pour certaines espèces d'hôtes. Ceci lui assure de multiples niches écologiques nécessaires pour sa survie, son maintien et sa dispersion dans l'environnement.

2.4. Faiblesses dans cette étude et comment nous avons tenu en compte

Dans cette étude, plusieurs faiblesses apparaissent qui sont dues notamment à la nature de l'échantillonnage par commodité. Comme la plupart des données écologiques (Etikan et al., 2016), nos données font l'objet d'un échantillonnage de convenance, dont l'utilisation est motivée par la disponibilité des données ayant une grande précision géographique (coordonnées GPS) en comparaison avec des données de surveillance passive.

2.4.1. Taille d'échantillon et puissance des associations

Malheureusement, comme dans la plupart des études de terrain qui sont menées dans un contexte d'émergence et dans des zones d'invasion, force est de constater que la prévalence des échantillons de tiques infectées est souvent très faible (Ogden et al., 2010). Il y a donc moins de données disponibles pour les analyses statistiques. Un effort considérable a été réalisé par l'ASPC pour essayer d'obtenir un maximum de données (p.ex. l'élargissement de la période des collectes de tiques, des zones échantillonnées et des sources telles que la végétation et les rongeurs). Cet effort pourrait être poursuit par l'investigation de nouvelles zones à partir du Centre jusqu'au Sud du Canada et par davantage de moyens pour échantillonner les régions où il y a des sites qui abritent une diversité importante, comme ceux identifiés dans cette étude.

2.4.2. Incertitude sur le type d'hôte

65,5% des échantillons d'ADN de *B. burgdorferi* utilisés pour les analyses de la dispersion des souches proviennent de tiques immatures (26 nymphes et 12 larves), qui sont connues pour parasiter de préférence les petits mammifères (Daniels & Fish, 1995; Walls et al., 1997). Par ailleurs, 34.5% des tiques sont des adultes (7 tiques collectées sur des cerfs et 13 autres collectées par la flanelle). Si l'hôte est le dénominateur commun entre toutes ces tiques (c'est-à-dire qu'elles se font déplacer par des hôtes), il reste qu'il existe une incertitude quant à l'espèce de l'hôte (au moins pour les tiques adultes). Ceci peut conduire à une erreur de type II, c'est-à-dire retenir H_0 qui présume l'absence de l'effet de la connectivité sur la probabilité que deux échantillons de deux sites portent un même ST quand elle est fausse. À titre d'exemple, si toutes les tiques sont portées par un hôte non affecté par l'impédance (résistance de surface) tel que le cerf, on peut s'attendre à ce qu'il n'y ait pas d'association.

2.4.3. Incertitude sur l'indice de résistance de surface

Baser les analyses sur des valeurs inadéquates de l'impédance (Cushman et al., 2013) entraînera des calculs incorrects de la connectivité du paysage. Pour limiter l'incertitude dans la résistance du paysage, nous avons utilisé ceux de la souris à pattes-blanches fournies par Marrotte et al. (2014) et Simon et al. (2014). En effet, ces données sont pertinentes, car les auteurs ont utilisé des données génétiques de ce rongeur pour calculer les valeurs de l'indice

de résistance de surface (Zeller et al., 2012). Ces valeurs ont été d'ailleurs validées pour le paysage de la Montérégie avec le modèle de dispersion des souris (Simon et al., 2014).

2.4.4. Les biais

a) Biais de classification

La carte de la classification du paysage réalisée pour cette étude était la deuxième source importante de données dans le calcul de la connectivité du paysage. La classification erronée est un biais commun dans des études d'évaluation du couvert végétal (Kleindl et al., 2015). Pour limiter ce biais, un effort considérable a été réalisé pour produire une classification précise en exploitant 9 images Landsat 8 de différentes saisons et années. La fusion des images panchromatiques apporte plus de précision. Ainsi, elle permet de travailler avec une résolution de 15m plutôt qu'une résolution de 30m. De même, l'exploitation de 3 images de 3 saisons sur 3 années a permis de minimiser les erreurs de mauvaise classification.

b) Biais de sélection

La nature de l'échantillonnage par commodité résultant des différentes campagnes de surveillance active qui ont été menées au fil des années est source de biais, notamment de biais de sélection qui limite la représentativité de la population générale de tiques *I. scapularis* du Canada. Les données sur les souches de *B. burgdorferi* pour cette étude sont disponibles pour un total de 73 visites effectuées durant 10 ans (2001, 2005, 2006, 2007, 2008, 2009, 2010, 2011, 2012 et 2013). Ces visites concernent 43 sites répartis sur un territoire du Canada qui couvre le Manitoba, l'Ontario, le Québec et les Maritimes. Par ailleurs, la répartition des visites et des sites par province et par année n'est pas équitable et non randomisée. Ceci a engendré une fluctuation d'effort d'échantillonnage entre les régions et les années.

c) Biais d'échantillonnage

Également, deux méthodes d'échantillonnage ont été utilisées : la flanelle et le piégeage des rongeurs qui ne sont pas effectuées ensembles systématiquement par site et par visite. Ceci a causé un biais d'échantillonnage car les tiques (incluant celles récupérées sur des cerfs chassés) ne sont pas prélevées toujours de la même façon sur tous les sites et durant toutes les visites. Cependant, la difficulté de randomiser à cette échelle géographique et

temporelle rend nécessaire l'échantillonnage par commodité, approche qui est statistiquement pertinente si on évite d'extrapoler les résultats en dehors du contexte de la zone d'étude (Etikan et al., 2016).

d) Biais de zone d'étude

Les régions n'ont pas été échantillonnées aux mêmes moments, ce qui rend difficile la comparaison entres elles, d'autant plus que la temporalité n'est pas respectée. Par ailleurs, l'aspect temporel dans le cas de *B. burgdorferi* ne peut pas être investigué avec les gènes de ménage sur un espace de 10 ans, car leur évolution est très lente et se fait sur des centaines, voire des milliers d'années (Margos et al. 2008). Si la zone d'étude s'élargit au fur et à mesure, il reste qu'elle fait partie d'un large territoire qui abrite des populations de tiques *I. scapularis* établies (Ogden et al., 2009) et où des cas humains ont été signalés (rapport sur la Surveillance nationale de la maladie de Lyme au Canada en 2016 : canada.ca/fr/sante-publique/).

Il est important de souligner que la population des tiques étudiée concerne uniquement des tiques infectées par *B. burgdorferi* qui est le dénominateur, le nombre d'individu de tiques infectées avec une souche donnée étant le numérateur.

En revanche, il n'y a pas de comparaisons faites avec des groupes de tiques négatives, l'objectif étant l'étude de l'évolution de *B. burgdorferi* dans son milieu naturel (c'est-à-dire dans la tique et l'hôte), et non l'étude du vecteur en tant que tel.

Un effort considérable a été effectué tout au long des analyses pour limiter l'effet de ces biais. D'abord, nous avons choisi des méthodes d'analyses adaptées à ces situations et qui permettent de tenir compte des effets possibles de ces biais sur les résultats. Par exemple, nous avons utilisé des modèles de régression mixte avec effet aléatoire et des tests non-paramétriques pour analyser nos données. De même, nous avons appliqué des critères rigoureux dans l'admission des données tels que les critères d'inclusion (p.ex. allèles qui constituent un ST sont sans sites ambigus) et d'exclusion (p.ex. élimination des données qui montrent des infections mixtes évidentes pour éviter la confusion, élimination des répliques de ST sur un même individu de rongeur pour éliminer l'impact possible de l'autocorrélation).

e) Biais d'attrition

Ce biais est causé par l'exclusion des données montrant des coïnfections. En effet, les coïnfections sont importantes car elles peuvent être une source de variation génétique pour la pathogénicité de *B. burgdorferi* (p.ex. il y a une possibilité de transfert du matériel génétique entre les souches co-infectantes) (Kurtenbach et al., 2001; Wójcik-Fatla et al., 2016). Les ignorer sous-estimerait ainsi la diversité génétique réelle de *B. burgdorferi* (Walter et al., 2016). Par conséquent, seuls les échantillons qui montrent des infections mixtes évidentes ne sont pas pris en compte délibérément dans cette étude. En effet, les méthodes moléculaires utilisées dans cette étude ne permettent pas d'identifier les souches dans les échantillons infectés avec plus qu'une souche, si une des souches n'est pas dominante (voir informations supplémentaires pour Mechai et al. 2015).

Par ailleurs, un reviewer de Mechai et al. (2015) a fait l'hypothèse que les souches trouvées dans cette étude en utilisant la méthode MLST sont des erreurs d'amplification et que les résultats étaient erronés. Pour répondre, le Laboratoire National de Microbiologie a mis en place une expérience qui a démontré que les allèles provenant des souches co-infectantes ne sont pas amplifiés au hasard dans chaque PCR (c'est-à-dire la souche présente à une concentration plus élevée sera préférentiellement amplifiée). Par conséquent, les nouveaux STs trouvés dans nos échantillons ne sont pas des artéfacts. Ces résultats ont été rapportés dans le matériel supplémentaire de Mechai et al. (2015 : voir Supplementary File 2). L'analyse des séquences à partir des traces d'ADN reste donc une méthode fiable pour identifier les infections mixtes tant que les souches mixtes sont dans des concentrations largement similaires. Par contre, lorsqu'une souche domine en termes du nombre de copies, elle produira les pics les plus clairs dans les traces d'ADN séquencées.

Il arrive fréquemment que le vecteur et l'hôte soient co-infectés par plusieurs génotypes de *B. burgdorferi* (Crowder et al., 2010). Certains génotypes arrivent mieux à se développer et à survivre dans le tube digestif de la tique et dans l'organisme de l'hôte (Kurtenbach et al., 2002). Le sang peut ainsi jouer un rôle de filtre favorisant l'un ou l'autre génotype (Kurtenbach et al., 2002; Anguita et al., 2003). Ce phénomène est aussi rapporté chez d'autres bactéries telles que la Bartonella chez les rongeurs et les puces (Gutierrez et al., 2014). 100% des tiques qui se co-nourrissent et qui figurent dans nos données sont toujours

infectées par la même souche. Ceci est dû à la méthodologie d'identification de la souche dominante indépendamment de la source (c'est-à-dire tique ou hôte). À titre d'exemple : 2 tiques (larve et nymphe ayant respectivement ID 33 et 36) se nourrissant sur un même individu de la souris à pattes-blanche (ID : 08-B-19-07-03) portent le même ST3, 3 nymphes (ID : GM11-039, GM11-041, GM11-042) se nourrissant sur un même individu du Campagnol à dos roux de Gapper (ID : GM-RBV1) portent le même ST1, et 4 nymphes (ID 125, 128, 129, 132) qui se nourrissent sur un même individu du tamia rayé (ID : 08-B-08-07-03) montrent un même ST14.

Pour améliorer la spécificité de cette approche et accroître la chance de détecter des souches rares, nous avons augmenté la taille d'échantillon de tiques testées, varié les sources (environnement et/ou hôtes) et élargit la zone d'étude. Dans l'avenir, l'utilisation des méthodes se basant sur le séquençage du génome complet est souhaitable, car ces méthodes ont une meilleure capacité d'identifier les souches dans les coïnfections (p.ex. comme dans Carpi et al., 2015).

Sur un total de 437 échantillons de tiques testées positives à l'ADN de *B. burgdorferi*, 273 ont été séquencées avec succès par la technique MLST (pour les 8 gènes) et 34 autres rejetés, car ils se sont révélés comme étant des infections mixtes ; 240 ont été séquencés avec succès pour le marqueur ospC et 25 autres échantillons montrant des infections mixtes ont été rejetés ; 258 ont été séquencés avec succès pour le spacer IGS et 30 autres échantillons rejetés en raison d'infections mixtes.

3. Les limites de l'étude

Certains des résultats de cette étude ne peuvent probablement être valables que dans des conditions similaires, que nous présentons comme suit :

- La zone d'étude est une zone d'émergence. Dans une zone fortement endémique, les choses peuvent être différentes (p.ex. l'effet de la dispersion sur la structure des souches peut être faible comparativement à l'effet d'autres facteurs comme l'immunité de l'hôte, la diversité des hôtes, la fragmentation du paysage).
- C'est une étude observationnelle basée sur un échantillonnage non-systématique et sur une taille d'échantillon relativement faible. Ceci a montré des associations qui nécessitent d'être confirmées par d'autres études.

- Le modèle de dispersion des souches présenté ici est dépendant du contexte d'émergence de la maladie de Lyme.
- Ce modèle a été réalisé dans le contexte du paysage de la Montérégie qui est très fragmenté.

IV. Perspectives d'avenir

1. Impact de ces résultats sur l'action en santé publique

1.1. Importance pour comprendre l'épidémiologie de la maladie de Lyme

Le paysage de la maladie de Lyme au Canada est constitué de 84 souches (30 nouvelles souches canadiennes et 54 autres sont originaires des ÉU). L'utilisation des données des ÉU et du Canada dans la phylogénie a permis de décrire une image globale de la structure génétique de *B. burgdorferi* à l'échelle continentale en Amérique du Nord. Ceci a révélé, en plus des liens de parenté évidents entre les souches des deux pays, que certaines d'entre elles n'ont été identifiées que dans l'un des pays, ce qui souligne l'importance de considérer les patrons régionaux et locaux pour comprendre la dynamique de la maladie de Lyme à différentes échelles.

Les politiques de santé publique concernant la maladie de Lyme au Canada doivent ainsi prendre en considération, dans leurs stratégies de lutte et de prévention, la dynamique locale et régionale de l'agent pathogène (en plus de l'aspect continental avec l'arrivée des oiseaux migrateurs). Trois régions ont abrité possiblement des populations refuges en commun avec le Midwest et le Nord-est des ÉU lors des périodes glaciaires : les Maritimes et le Québec, le Long point en Ontario, le Nord-ouest d'Ontario et le Sud de Manitoba. Par conséquent, elles font face à une diversité génétique de *B. burgdorferi* spécifique qui peut engendrer des symptômes différents, ce qui peut nécessiter des politiques adaptées à chaque situation.

1.2. Importance pour l'évaluation du risque

La variation régionale des souches (diversité et fréquence) révélée dans cette étude peut être un indicateur de la variation régionale de la pathogénicité et de la probable faiblesse des tests diagnostiques rapportée par Ogden et al. (2017). La variation locale des souches due aux communautés d'hôtes peut être un indicateur de la pathogénicité différentielle des souches à l'échelle locale, ce qui pourrait donner une symptomatologie spécifique et différente selon

les régions. De plus, le patron de dispersion des souches dépendant du paysage, le risque pourrait varier selon la zone géographique.

Par conséquent, dans le futur il est important de développer des modèles de risque et des cartes de risque pour : i) le risque de la maladie de Lyme, ii) le risque des souches hautement pathogènes, iii) le risque de diagnostiquer des faux-négatifs de la maladie de Lyme, iv) le risque de la dispersion de la maladie de Lyme pour les années à venir.

1.3.Amélioration du système de surveillance de la maladie de Lyme au Canada

Les résultats de cette étude suggèrent d'élargir le système de surveillance aux différentes souches de *B. burgdorferi* et non seulement aux cas humains et au vecteur. Ceci montre des possibilités de concevoir des surveillances intelligentes à l'échelle locale qui tiennent en compte les connaissances sur l'écologie de *B. burgdorferi*. En effet, le mouvement des hôtes est nécessaire pour pouvoir prédire la trajectoire de chaque souche afin de mieux identifier les populations humaines à risque et de cibler les politiques de prévention en santé publique. Ainsi, la prédiction du lieu et de la vitesse de l'occurrence des futurs cas humains devient possible dans un contexte d'émergence.

2. Axes de recherches

À l'issu de notre travail, différentes possibilités d'axes de recherche s'ouvrent : i) la diversité génétique de *B. burgdorferi* au Canada pourrait être investiguée de plus près *in vitro* et *in vivo* sur des modèles de souris et de tamia par des cultures bactériennes (p.ex. pour créer des souches pures) pour mesurer l'effet sur la pathogénicité différentielle *in fine* sur des cas humains, ii) tester les kits d'identification actuels sur des souches pures connues au laboratoire et développer (s'il y a lieu) des méthodes de diagnostic plus spécifiques en exploitant le génome complet de la bactérie, iii) investiguer et mesurer l'importance des associations hôtes-souches dans le processus d'évaluation du risque de la maladie de Lyme à l'échelle régionale et nationale au Canada (notamment en produisant des cartes de risque associées aux différents symptômes présents au Canada), iv) investiguer les causes des associations hôtes-souches (avec des études de terrain plus vastes et des études de transmission au laboratoire) afin de détecter des mécanismes de transmissibilité et d'infection différentielle que pourraient cacher ces associations (p.ex. la recherche d'autres protéines de surface), v) incorporer le patron de

dispersion des souches dans les modèles de risque (le modèle proposé par Ogden et al. (2008b) utilise seulement la diffusion brute de *B. burgdorferi*), vi) continuer les études phylogénétiques et phylogéographiques pour mieux suivre l'évolution de *B. burgdorferi* au Canada.

V. Conclusion générale

Nous avons produit le premier catalogue complet (mais non exhaustif) de souches de *B. burgdorferi* au Canada en utilisant les méthodes de multi-locus les plus performantes pour le génotypage de cette bactérie. Ceci a révélé pour la première fois l'étendue de la diversité génétique de *B. burgdorferi* au Canada, qui est plus importante dans le Sud-ouest Canadien. Sa diversité au Canada est comparable à sa diversité aux ÉU indiquant que l'épidémiologie de la maladie de Lyme peut prendre l'ampleur qu'on lui connait actuellement au Nord des ÉU. Cette diversité peut avoir des implications directes sur le diagnostic de la maladie de Lyme en Amérique du Nord, mais aussi sur la pathogénicité des différentes souches de *B. burgdorferi*.

L'exploration de l'histoire évolutive de *B. burgdorferi* a permis de produire une première image phylogéographique nécessaire pour comprendre l'épidémiologie évolutive de la maladie de Lyme au Canada. Par conséquent, cette étude fournit un premier modèle phylogéographique de *B. burgdorferi* qui associe d'anciens changements climatiques (périodes glaciaires-interglaciaires) et des changements actuels (réchauffement climatique) avec un patron écologique pour expliquer la dispersion passée, contemporaine, mais aussi future de *B. burgdorferi* en Amérique du Nord. Les modèles de prédiction doivent prendre en considération l'effet de ce patron sur l'épidémiologie de la maladie de Lyme. Des recherches supplémentaires sont nécessaires pour mieux décrire l'étendue et la force des associations hôte-génotype. Pour pouvoir prédire leurs conséquences sur la santé humaine, il est important de mieux comprendre comment elles se produisent mécaniquement, possiblement à l'aide des recherches expérimentales sur des modèles de souris et de tamia. De même, la recherche d'autres protéines de surface pourra mettre à jour le mécanisme d'adaptation qu'utilise *B. burgdorferi* pour survivre au système immunitaire de l'organisme. Ceci, peut être conduit notamment en exploitant le génome complet de cette bactérie.

Bibliographie

Aguero-Rosenfeld, M. E., Wang, G., Schwartz, I., & Wormser, G. P. (2005). Diagnosis of Lyme Borreliosis. *Clin Microbiol Rev, 18*(3), 484-509.

Alarcón, d. N. B., Díaz-Bello, Z., Colmenares, C., Ruiz-Guevara, R., Mauriello, L., Zavala-Jaspe, R., & Paiva, M. (2010). Large urban outbreak of orally acquired acute Chagas disease at a school in Caracas, Venezuela. *J Infect Dis, 201*(9), 1308-1315.

Alekseev, A. N., Dubinina, H. V., Van De Pol, I., & Schouls, L. M. (2001). Identification of Ehrlichia spp. and *Borrelia burgdorferi* in Ixodes ticks in the Baltic regions of Russia. *J Clin Microbiol, 39(6),* 2237-2242.

Allan, B. F., Keesing, F., & Ostfeld, R. S. (2003). Effect of forest fragmentation on Lyme disease risk. *Conservation Biology, 17*(1), 267-272.

Allen, H. B., Vin, H., Warner, C., & Joshi, S. (2016). Lyme Disease: Beyond Erythema Migrans. *Journal of Clinical & Experimental Dermatology Research, 2016*.

Amos, J. N., Bennett, A. F., Mac Nally, R., Newell, G., Pavlova, A., Radford, J. Q., Thomson, J., White, M., Sunnucks, P. (2012). Predicting landscape-genetic consequences of habitat loss, fragmentation and mobility for multiple species of woodland birds. *PLoS One, 7*(2), e30888.

Anderson JM and Norris DE. (2006). Genetic Diversity of *Borrelia burgdorferi* Sensu Stricto in *Peromyscus leucopus*, the Primary Reservoir of Lyme Disease in a Region of Endemicity in Southern Maryland. *Appl. Environ. Microbiol.*, *72*(8): 5331–5341.

Ang, C. W., Notermans, D. W., Hommes, M., Simoons-Smit, A. M., & Herremans, T. (2011). Large differences between test strategies for the detection of anti-Borrelia antibodies are revealed by comparing eight ELISAs and five immunoblots. *European journal of clinical microbiology & infectious diseases, 30(8),* 1027-1032.

Anguita, J., Hedrick, M. N., & Fikrig, E. (2003). Adaptation of *Borrelia burgdorferi* in the tick and the mammalian host. *FEMS microbiology reviews, 27*(4): 493-504.

Appel, M. J. (1990). Lyme disease in dogs and cats. *Compendium on continuing education for the practicing veterinarian, 12*(5), 617-626.

Bacon, R. M., Kugeler, K. J., & Mead, P. S. (2008). Surveillance for Lyme disease--United States, 1992-2006. Atlanta, GA: Department of Health & Human Services, Centers for Disease Control and Prevention.

Barbour A, Garon CF. (1988). The genes encoding major surface proteins of *Borrelia burgdorferi* are located on a plasmid. *Annals of the New York Academie of Sciences*, 539:144–153.

Barbour, A. G., & Fish, D. (1993). The biological and social phenomenon of Lyme disease. *SCIENCE-NEW YORK THEN WASHINGTON-, 260*, 1610-1610.

Barker, I. K., McEwen, S. A., Surgeoner, G. A., & Artsob, H. (1988). *Borrelia burgdorferi*, the agent of Lyme disease, in tick vectors and wildlife reservoirs in southern Ontario. Ont. *Dis. Surv. Rep, 9,* 151-154.

Becker, N. S., Margos, G., Blum, H., Krebs, S., Graf, A., Lane, R. S., & Fingerle, V. (2016). Recurrent evolution of host and vector association in bacteria of the Borrelia burgdorferi sensu lato species complex. *BMC Genomics, 17*(1), 734.

Bennett, A. F., Henein, K., & Merriam, G. (1994). Corridor use and the elements of corridor quality: Chipmunks and fencerows in a farmland mosaic. *Biological Conservation, 68*(2), 155-165.

Bennett, K. D. (1997). *Evolution and ecology: the pace of life*: Cambridge University Press.

Berdoy, M., Webster, J. P., & Macdonald, D. W. (2000). Fatal attraction in rats infected with *Toxoplasma gondii*. *Proceedings of the Royal Society of London B: Biological Sciences, 267(1452)*, 1591-1594.

Bockenstedt, L. K., Liu, N., Schwartz, I., & Fish, D. (2006). MyD88 deficiency enhances acquisition and transmission of *Borrelia burgdorferi* by Ixodes scapularis ticks. *Infect Immun, 74(4)*, 2154-2160.

Bouchard, C., Leonard, E., Koffi, J. K., Pelcat, Y., Peregrine, A., Chilton, N., Rochon, K., Lysyk, T., Lindsay, L. R., Ogden, N. H. (2015). The increasing risk of Lyme disease in Canada. *Can Vet J, 56(7),* 693-699.

Bown, K. J., Begon, M., Bennett, M., Birtles, R. J., Burthe, S., Lambin, X., Telfer, S., Woldehiwet, Z., & Ogden, N. H. (2006). Sympatric *Ixodes trianguliceps* and *Ixodes ricinus* ticks feeding on field voles (*Microtus agrestis*): potential for increased risk of

Anaplasma phagocytophilum in the United Kingdom?. *Vector-Borne & Zoonotic Diseases, 6(4)*, 404-410.

Brown, E. L., Guo, B. P., O'Neal, P., & Hook, M. (1999). Adherence of *Borrelia burgdorferi*. Identification of critical lysine residues in DbpA required for decorin binding. *J Biol Chem, 274(37),* 26272-26278.

Boyle, S. A., Kennedy, C. M., Torres, J., Colman, K., Pérez-Estigarribia, P. E., & Noé, U. (2014). High-resolution satellite imagery is an important yet underutilized resource in conservation biology. *PLoS One, 9*(1), e86908.

Bowne, D. R., & Bowers, M. A. (2004). Interpatch movements in spatially structured populations: a literature review. *Landscape Ecology, 19*(1), 1-20.

Brownstein, J. S., Holford,T. R., & Fish, D. A. (2003). climate-based model predicts the spatial distribution of the Lyme disease vector Ixodes scapularis in the United States. Environ Health Perspect. 111(9):1152-7.

Brisson, D., & Dykhuizen, D. E. (2004). ospC diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics, 168*(2), 713-722. doi:10.1534/genetics.104.028738

Brisson, D., Drecktrah, D., Eggers, C. H., & Samuels, D. S. (2012). Genetics of *Borrelia burgdorferi*. *Annual review of genetics, 46*: 515-536.

Broquet, T., Ray, N., Petit, E., Fryxell, J. M., & Burel, F. (2006). Genetic isolation by distance and landscape connectivity in the American marten (*Martes americana*). *Landscape Ecology, 21*(6), 877-889.

Brownstein, J. S., Holford, T. R., & Fish, D. (2005). Effect of climate change on Lyme disease risk in North America. *EcoHealth*, 2(1), 38-46.

Bunikis, J., Garpmo, U., Tsao, J., Berglund, J., Fish, D., & Barbour, A. G. (2004). Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology, 150(Pt 6)*, 1741-1755.

Burgdorfer, W., Barbour, A. G., Hayes, S. F., Benach, J. L., Grunwaldt, E., & Davis, J. P. (1982). Lyme disease-a tick-borne spirochetosis? *Science, 216*(4552), 1317-1319.

Carpi, G., Walter, K. S., Bent, S. J., Hoen, A. G., Diuk-Wasser, M., & Caccone, A. (2015). Whole genome capture of vector-borne pathogens from mixed DNA samples: a case study of *Borrelia burgdorferi*. *BMC genomics, 16(1),* 434.

Castillo-Ramirez, S., Fingerle, V., Jungnick, S., Straubinger, R. K., Krebs, S., Blum, H., Meinel, D. M., Hofmann, H., Guertler, P., Sing, A., Margos, G. (2016). Trans-Atlantic exchanges have shaped the population structure of the Lyme disease agent *Borrelia burgdorferi* sensu stricto. *Sci Rep, 6,* 22794.

Cheng, L., Connor, T. R., Aanensen, D. M., Spratt, B. G., & Corander, J. (2011). Bayesian semi-supervised classification of bacterial samples using MLST databases. *BMC bioinformatics, 12*(1): 302.

Clauzel, C., Foltête, J.-C., Girardet, X., & Vuidel, G. (2016). Graphab 2.0 Reference Manual Command Line Interface.

Combes, C. (1997). Fitness of parasites: pathology and selection. *International journal for parasitology, 27(1),* 1-10.

Crida, A., & Manel, S. (2007). WOMBSOFT: an R package that implements the Wombling method to identify genetic boundary. *Molecular Ecology Resources, 7*(4), 588-591.

Crowder, C. D., Matthews, H. E., Schutzer, S., Rounds, M. A., Luft, B. J., Nolte, O., . . . Sampath, R. (2010). Genotypic variation and mixtures of Lyme Borrelia in Ixodes ticks from North America and Europe. *PLoS One, 5*(5), e10650.

Cumming, G. S., & Guégan, J. F. (2006). Food webs and disease: is pathogen diversity limited by vector diversity?. *Ecohealth, 3(3),* 163-170.

Daniels, T. J., & Fish, D. (1995). Effect of deer exclusion on the abundance of immature Ixodes scapularis (Acari: Ixodidae) parasitizing small and medium-sized mammals. *Journal of Medical Entomology, 32*(1), 5-11.

de la Fuente, J., Estrada-Pena, A., Venzal, J. M., Kocan, K. M., & Sonenshine, D. E. (2008). Overview: Ticks as vectors of pathogens that cause disease in humans and animals. *Front Biosci, 13*, 6938-6946.

Derdakova, M., Dudioak, V., Brei, B., Brownstein, J. S., Schwartz, I., & Fish, D. (2004). Interaction and transmission of two *Borrelia burgdorferi* sensu stricto strains in a tick-rodent maintenance system. *Appl Environ Microbiol, 70*(11), 6783-6788.

Devevey, G., Dang, T., Graves, C. J., Murray, S., & Brisson, D. 2015. First arrived takes all: inhibitory priority effects dominate competition between co-infecting *Borrelia burgdorferi* strains. *BMC Microbiol, 2015*, 15: 61.

Divers, T., Chang, Y., Jacobson, R., & McDonough, S. (2001). Lyme disease in horses. *compendium on continuing education for the practising veterinarian-north american edition-, 23(4),* 375-382.

Diuk-Wasser, M. A., Hoen, A. G., Cislo, P., Brinkerhoff, R., Hamer, S. A., Rowland, M., Cortinas, R., Vourc'h, G., Melton, F., Hickling, G. J., Tsao, J. I., Bunikis, J., Barbour, A. G., Kitron, U., Piesman, J., Fish, D. (2012). Human risk of infection with *Borrelia burgdorferi*, the Lyme disease agent, in eastern United States. *Am J Trop Med Hyg, 86(2)*, 320-327.

Dragoo, J. W., Lackey, J. A., Moore, K. E., Lessa, E. P., Cook, J. A., & Yates, T. L. (2006). Phylogeography of the deer mouse (*Peromyscus maniculatus*) provides a predictive framework for research on hantaviruses. *J Gen Virol, 87(Pt 7)*, 1997-2003.

Dykhuizen, D. E., Brisson, D., Sandigursky, S., Wormser, G. P., Nowakowski, J., Nadelman, R. B., & Schwartz, I. (2008). The propensity of different *Borrelia burgdorferi* sensu stricto genotypes to cause disseminated infections in humans. *Am J Trop Med Hyg, 78(5)*, 806-810.

Embers, M. E., Ramamoorthy, R., & Philipp, M. T. (2004). Survival strategies of *Borrelia burgdorferi*, the etiologic agent of Lyme disease. Microbes Infect, 6(3), 312-318.

Estrada-Peña, A., & de la Fuente, J. (2014). The ecology of ticks and epidemiology of tick-borne viral diseases. *Antiviral research, 108*, 104-128.

Estrada-Peña, A., Sprong, H., Cabezas-Cruz, A., de la Fuente, J., Ramo, A., & Coipan, E. C. (2016). Nested coevolutionary networks shape the ecological relationships of ticks, hosts, and the Lyme disease bacteria of the *Borrelia burgdorferi* (sl) complex. *Parasit Vectors, 9*(1), 517.

Etikan, I., Musa, S. A., & Alkassim, R. S. (2016). Comparison of Convenience Sampling and Purposive Sampling. *American Journal of Theoretical and Applied Statistics, 5*(1), 1-4.

Fauci, A. S. (2005). Emerging and Reemerging Infectious Diseases: The Perpetual Challenge. *Academic Medicine, 80*(12), 1079-1085.

Fergus, C., & Hansen, A. (2000). *Wildlife of Pennsylvania and the Northeast*: Stackpole Books.

Fialkowski, V. A., Foster, E., Signs, K., & Stobierski, M. G. (2016). One Health in Action: Lyme Disease. *Online Journal of Public Health Informatics, 8*(1).

Fiset, J., Tessier, N., Millien, V., & Lapointe, F. J. (2015). Phylogeographic Structure of the White-Footed Mouse and the Deer Mouse, Two Lyme Disease Reservoir Hosts in Quebec. *PLoS One, 10*(12), e0144112.

Floden, A. M., Watt, J. A., & Brissette, C. A. (2011). *Borrelia burgdorferi* enolase is a surface-exposed plasminogen binding protein. *PLoS One, 6(11),* e27502.

Fukunaga M, Okada K, Nakao M, Konishi T, Sato Y. (1996). Phylogenetic analysis of *Borrelia* species based on flagellin gene sequences and its application for molecular typing of Lyme disease *borreliae*. *Int J Syst Bacteriol*, 46:898–905.

Gatewood, A. G., Liebman, K. A., Vourc'h, G., Bunikis, J., Hamer, S. A., Cortinas, R., Melton, F., Cislo, P., Kitron, U., Tsao, J., Barbour, A. G., Fish, D., Diuk-Wasser, M. A. (2009). Climate and tick seasonality are predictors of *Borrelia burgdorferi* genotype distribution. *Appl Environ Microbiol, 75*(8), 2476-2483.

Gęca, A., Mazurek, U., Muc-Wierzgoń, M., Nowakowska-Zajdel, E., Niedworok, E., Ziółko, E., & Kokot, T. (2016). Znaczenie czynnika H w patogenezie zakażenia krętkami Borrelia The role of complement factor H in the pathogenesis of *Borrelia* infection. *Journal cover, 70.*

Gebbia, J. A., Coleman, J. L., & Benach, J. L. (2001). *Borrelia* spirochetes upregulate release and activation of matrix metalloproteinase gelatinase B (MMP-9) and collagenase 1 (MMP-1) in human cells. *Infection and immunity, 69(1),* 456-462.

Gern L. (2008). *Borrelia burgdorferi* sensu lato, the agent of Lyme borreliosis: life in the wilds. *Parasite*, 15: 244–247.

Ginsberg, H. S., Rulison, E. L., Azevedo, A., Pang, G. C., Kuczaj, I. M., Tsao, J. I., & LeBrun, R. A. (2014). Comparison of survival patterns of northern and southern genotypes of the North American tick *Ixodes scapularis* (Acari: *Ixodidae*) under northern and southern conditions. *Parasit Vectors, 7,* 394.

Grimm, D., Eggers, C. H., Caimano, M. J., Tilly, K., Stewart, P. E., Elias, A. F., Radolf, J. D., Rosa, P. A. (2004). Experimental assessment of the roles of linear plasmids lp25 and

lp28-1 of *Borrelia burgdorferi* throughout the infectious cycle. *Infect Immun, 72(10),* 5938-5946.

Gubler, D. J., Reiter, P., Ebi, K. L., Yap, W., Nasci, R., & Patz, J. A. (2001). Climate variability and change in the United States: potential impacts on vector-and rodent-borne diseases. *Environmental health perspectives, 109(Suppl 2),* 223.

Guerra M, Walker E, Jones C, Paskewitz S, Cortinas MR, Stancil A, Beck L, Bobo M, Kitron U. (2002). Predicting the risk of Lyme disease: habitat suitability for Ixodes scapularis in the north central United States. *Emerg Infect Dis.* 8(3), 289-97.

Gutierrez, R., Morick, D., Cohen, C., Hawlena, H., & Harrus, S. (2014). The effect of ecological and temporal factors on the composition of *Bartonella* infection in rodents and their fleas. *ISME J, 8*(8), 1598-1608.

Hacker, J., & Kaper, J. B. (2000). Pathogenicity islands and the evolution of microbes. *Annual Reviews in Microbiology, 54*(1), 641-679.

Hadly, E. A., Ramakrishnan, U. M. A., Chan, Y. L., Van Tuinen, M., O'Keefe, K., Spaeth, P. A., & Conroy, C. J. (2004). Genetic response to climatic change: insights from ancient DNA and phylochronology. *PLoS Biol, 2(10),* e290.

Hanincová, K., Kurtenbach, K., Diuk-Wasser, M., Brei, B., & Fish, D. (2006). Epidemic spread of Lyme borreliosis, northeastern United States. *Emerg Infect Dis,* 12.

Hanincova, K., Mukherjee, P., Ogden, N. H., Margos, G., Wormser, G. P., Reed, K. D., Meece, J. K., Vandermause, M. F., Schwartz, I. (2013). Multilocus sequence typing of *Borrelia burgdorferi* suggests existence of lineages with differential pathogenic properties in humans. *PLoS One, 8*(9), e73066.

Hanincová, K., Liveris, D., Sandigursky, S., Wormser, G. P., & Schwartz, I. (2008a). *Borrelia burgdorferi* sensu stricto is clonal in patients with early Lyme borreliosis. *Applied and environmental microbiology, 74*(16), 5008-5014.

Hanincova, K., Ogden, N. H., Diuk-Wasser, M., Pappas, C. J., Iyer, R., Fish, D., Schwartz, I., & Kurtenbach, K. (2008b). Fitness variation of *Borrelia burgdorferi* sensu stricto strains in mice. *Appl Environ Microbiol, 74*(1), 153-157.

Harris, M., & Reza, J. (2012). *Global report for research on infectious diseases of poverty*: World Health Organization.

Hartemink, N. A., Randolph, S. E., Davis, S. A., & Heesterbeek, J. A. P. (2008). The basic reproduction number for complex disease systems: Defining R 0 for tick-borne infections. *The American Naturalist*, *171*(6), 743-754.

Hoon-Hanks, L. L., Morton, E. A., Lybecker, M. C., Battisti, J. M., Samuels, D. S., & Drecktrah, D. (2012). *Borrelia burgdorferi* malQ mutants utilize disaccharides and traverse the enzootic cycle. *FEMS Immunol Med Microbiol, 66(2),* 157-165.

Hoen, A. G., Margos, G., Bent, S. J., Diuk-Wasser, M. A., Barbour, A., Kurtenbach, K., & Fish, D. (2009). Phylogeography of *Borrelia burgdorferi* in the eastern United States reflects multiple independent Lyme disease emergence events. *Proc Natl Acad Sci U S A, 106*(35), 15013-15018.

Holdenrieder, O., Pautasso, M., Weisberg, P. J., & Lonsdale, D. (2004). Tree diseases and landscape processes: the challenge of landscape pathology. *Trends in Ecology & Evolution, 19*(8), 446-452.

Houwerzyl, J., Root, J., & Hoogkamp-Korstanje, J. (1984). A case of Lyme disease with cardiac involvement in the Netherlands. *Infection, 12*(5), 358-358.

Huelsenbeck, J. P., Larget, B., & Alfaro, M. E. (2004). Bayesian phylogenetic model selection using reversible jump Markov chain Monte Carlo. *Mol Biol Evol, 21*(6), 1123-1133.

Humphrey, P. T., Caporale, D. A., & Brisson, D. (2010). Uncoordinated phylogeography of *Borrelia burgdorferi* and its tick vector, *Ixodes scapularis*. *Evolution, 64*(9), 2653-2663.

Imbens, G. W., & Kolesar, M. (2012). Robust standard errors in small samples: Some practical advice. *Review of Economics and Statistics*(0).

Indest, K. J., Ramamoorthy, R., Solé, M., Gilmore, R. D., Johnson, B. J., & Philipp, M. T. (1997). Cell-density-dependent expression of *Borrelia burgdorferi* lipoproteins in vitro. *Infection and immunity, 65(4),* 1165-1171.

Jacquot, M., Abrial, D., Gasqui, P., Bord, S., Marsot, M., Masseglia, S., Pion, A., Poux, V., Zilliox, L., Chapuis, J. L., Vourc'h, G., & Bailly X. (2016). Multiple independent transmission cycles of a tick-borne pathogen within a local host community. *Scientific reports,* 6, 31273.

Johnson, B. J. (2011). Laboratory diagnostic testing for *Borrelia burgdorferi* infection. Lyme Disease: An evidence-based approach. *CAB International,* 73-88.

Johnson, T.L., Bjork, J.K., Neitzel, D.F., Dorr, F.M., Schiffman, E.K., Eisen, R.J. (2016). Habitat Suitability Model for the Distribution of *Ixodes scapularis* (Acari: *Ixodidae*) in Minnesota. *J Med Entomol. 53(3),* 598-606.

Jones, K. E., Patel, N. G., Levy, M. A., Storeygard, A., Balk, D., Gittleman, J. L., & Daszak, P. (2008). Global trends in emerging infectious diseases. *Nature, 451*(7181), 990-993.

Jungnick, S., Margos, G., Rieger, M., Dzaferovic, E., Bent, S. J., Overzier, E., Silaghi, C., Walder, G., Wex, F., Koloczek, J., Sing, A., Fingerle, V. (2015). *Borrelia burgdorferi* sensu stricto and *Borrelia afzelii*: Population structure and differential pathogenicity. *Int J Med Microbiol, 305*(7), 673-681.

Kempf, F., Boulinier, T., De Meeûs, T., Arnathau, C., & McCOY, K. D. (2009). Recent evolution of host-associated divergence in the seabird tick *Ixodes uriae*. *Molecular Ecology, 18*(21), 4450-4462.

Klopfstein, S., Currat, M., & Excoffier, L. (2006). The fate of mutations surfing on the wave of a range expansion. Mol Biol Evol, 23(3), 482-490.

Koedel, U., & Pfister, H.-W. (2017). Lyme neuroborreliosis. *Current Opinion in Infectious Diseases, 30*(1), 101-107.

Koella, J. C., SÖrensen, F. L., & Anderson, R. A. (1998). The malaria parasite, *Plasmodium falciparum*, increases the frequency of multiple feeding of its mosquito vector, Anopheles gambiae. *Proceedings of the Royal Society of London B: Biological Sciences, 265(1398)*, 763-768.

Kozakiewicz, M. (1993). Habitat isolation and ecological barriers-the effect on small mammal populations and communities. *Acta Theriologica, 38*(1), 1-30.

Krause, P. J., Narasimhan, S., Wormser, G. P., Rollend, L., Fikrig, E., Lepore, T., Fish, D. (2013). Human *Borrelia miyamotoi* infection in the United States. *New England Journal of Medicine, 368*(3), 291-293.

Kuehn, B. M. (2013). CDC estimates 300 000 US cases of Lyme disease annually. *JAMA, 310*(11), 1110-1110.

Kurtenbach, K., De Michelis, S., Etti, S., Schafer, S. M., Sewell, H. S., Brade, V., & Kraiczy, P. (2002). Host association of *Borrelia burgdorferi* sensu lato-the key role of host complement. *Trends Microbiol, 10*(2), 74-79.

Kurtenbach, K., De Michelis, S., Sewell, H. S., Etti, S., Schafer, S. M., Hails, R., Collares-Pereira, M., Santos-Reis, M., Hanincova, K., Labuda, M., Bormane, A., Donaghy, M. (2001). Distinct combinations of *Borrelia burgdorferi* sensu lato genospecies found in individual questing ticks from Europe. *Appl Environ Microbiol, 67*(10), 4926-4929.

Kurtenbach, K., Hanincova, K., Tsao, J. I., Margos, G., Fish, D., & Ogden, N. H. (2006). Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nat Rev Microbiol, 4*(9), 660-669.

Lartillot, N. (2015). Probabilistic models of eukaryotic evolution: time for integration. *Philosophical Transactions of the Royal Society B: Biological Sciences, 370*(1678), 20140338.

Leighton, P. A., Koffi, J. K., Pelcat, Y., Lindsay, L. R., & Ogden, N. H. (2012). Predicting the speed of tick invasion: an empirical model of range expansion for the Lyme disease vector *Ixodes scapularis* in Canada. *Journal of Applied Ecology, 49(2), 457-464.*

Lin T, Oliver JH, Jr., Gao L. (2002). Genetic Diversity of the Outer Surface Protein C Gene of Southern *Borrelia* Isolates and Its Possible Epidemiological, Clinical, and Pathogenetic Implications. *J Clin Microbiol*, 40, 2572–2583.

Lindsay, L. R., Barker, I. K., Surgeoner, G. A., McEwen, S. A., & Campbell, G. D. (1997). Duration of *Borrelia burgdorferi* infectivity in white-footed mice for the tick vector *Ixodes scapularis* under laboratory and field conditions in Ontario. *J Wildl Dis, 33*(4), 766-775.

Liu, X. Y., & Bonnet, S. I. (2014). Hard tick factors implicated in pathogen transmission. *PLoS Negl Trop Dis, 8(1), e2566*.

Liveris D, Gazumyan A, Schwartz I. (1995). Molecular typing of *Borrelia burgdorferi* sensu lato by PCR-restriction fragment length polymorphism analysis. *J Clin Microbio*l, 33, 589–595.

Margos, G., Gatewood, A. G., Aanensen, D. M., Hanincová, K., Terekhova, D., Vollmer, S. A., Bormane, A. (2008). MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*. *Proceedings of the National Academy of Sciences, 105*(25), 8730-8735.

Margos, G., Hojgaard, A., Lane, R. S., Cornet, M., Fingerle, V., Rudenko, N., Ogden, N., Aanensen, D. M., Fish, D., Piesman, J. (2010). Multilocus sequence analysis of

Borrelia bissettii strains from North America reveals a new Borrelia species, *Borrelia kurtenbachii*. *Ticks Tick Borne Dis, 1*(4), 151-158.

Margos, G., Tsao, J. I., Castillo-Ramirez, S., Girard, Y. A., Hamer, S. A., Hoen, A. G., Lane, R. S., Raper, S. L., Ogden, N. H. (2012). Two boundaries separate *Borrelia burgdorferi* populations in North America. *Appl Environ Microbiol, 78*(17), 6059-6067.

Margos, G., Vollmer, S. A., Ogden, N. H., & Fish, D. (2011). Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato. *Infect Genet Evol, 11, 1545-1563*.

Marrotte, R. R., Gonzalez, A., & Millien, V. (2014). Landscape resistance and habitat combine to provide an optimal model of genetic structure and connectivity at the range margin of a small mammal. *Molecular Ecology, 23*(16), 3983-3998.

Marzluff, J., Shulenberger, E., Endlicher, W., Alberti, m., Bradley, G., Ryan, C., ZumBrunnen, C., Simon, U. (2008). *Urban Ecology: An International Perspective on the Interaction Between Humans and Nature*: Springer US.

Matuschka, F. R., & Spielman, A. (1986). The emergence of Lyme disease in a changing environment in North America and central Europe. *Exp Appl Acarol, 2*(4), 337-353.

McCoy, K. D., Boulinier, T., Tirard, C., & Michalakis, Y. (2003). Host-dependent genetic structure of parasite populations: differential dispersal of seabird tick host races. *Evolution (N Y), 57.*

McCoy, K. D., Léger, E., & Dietrich, M. (2013). Host specialization in ticks and transmission of tick-borne diseases: a review. *Frontiers in cellular and infection microbiology, 3.*

McRae, B. H., Dickson, B. G., Keitt, T. H., & Shah, V. B. (2008). Using circuit theory to model connectivity in ecology, evolution, and conservation. *Ecology, 89*(10), 2712-2724.

McRae, B. H., & Nürnberger, B. (2006). Isolation by resistance. *Evolution, 60*(8), 1551-1561.

Mead, P. S. (2015). Epidemiology of Lyme disease. *Infect Dis Clin North Am, 29*(2), 187-210.

Mechai, S., Feil, E. J., Gariepy, T. D., Gregory, T. R., Lindsay, L. R., Millien, V., & Ogden, N. H. (2013). Investigation of the population structure of the tick vector of Lyme disease *Ixodes scapularis* (Acari: *Ixodidae*) in Canada using mitochondrial cytochrome C oxidase subunit I gene sequences. *J Med Entomol, 50(3), 560-570.*

Mechai, S., Margos, G., Feil, E. J., Barairo, N., Lindsay, L. R., Michel, P., & Ogden, N. H. (2016). Evidence for Host-Genotype Associations of *Borrelia burgdorferi* Sensu Stricto. *PLoS One, 11*(2), e0149345.

Mechai, S., Margos, G., Feil, E. J., Lindsay, L. R., & Ogden, N. H. (2015). Complex population structure of *Borrelia burgdorferi* in southeastern and south central Canada as revealed by phylogeographic analysis. *Appl Environ Microbiol, 81(4), 1309-1318.*

Medlock, J. M., Hansford, K. M., Bormane, A., Derdakova, M., Estrada-Peña, A., George, J.-C., Golovljova, I., Jaenson, T.G., Jensen, J.K., Jensen, P.M., Kazimirova, M., Oteo, J.A., Papa, A., Pfister, K., Plantard, O., Randolph, S.E., Rizzoli, A., Santos-Silva, M. M., Sprong, H., Vial, L., Hendrickx, G., Zeller, H., Van Bortel, W. (2013). Driving forces for changes in geographical distribution of *Ixodes ricinus* ticks in Europe. *Parasit Vectors, 6*(1), 1.

Merriam, G., Kozakiewicz, M., Tsuchiya, E., & Hawley, K. (1989). Barriers as boundaries for metapopulations and demes of *Peromyscus leucopus* in farm landscapes. *Landscape Ecology, 2*(4), 227-235.

Merriam, G., & Lanoue, A. (1990). Corridor use by small mammals: field measurement for three experimental types of *Peromyscus leucopus*. *Landscape Ecology, 4*(2), 123-131.

Mongodin, E. F., Casjens, S. R., Bruno, J. F., Xu, Y., Drabek, E. F., Riley, D. R., Cantarel, B. L., Pagan, P. E., Hernandez, Y. A., Vargas, L. C., Dunn, J. J., Schutzer, S. E., Fraser, C. M., Qiu, W. G., Luft, B. J. (2013). Inter- and intra-specific pan-genomes of *Borrelia burgdorferi* sensu lato: genome stability and adaptive radiation. *BMC Genomics, 14*, 693.

Morlon, H., Kemps, B. D., Plotkin, J. B., & Brisson, D. (2012). explosive radiation of a bacterial species group. *Evolution, 66*(8), 10.1111/j.1558-5646.2012.01598.x.

Morshed, M. G., Scott, J. D., Fernando, K., Beati, L., Mazerolle, D. F., Geddes, G., & Durden, L. A. (2005). Migratory songbirds disperse ticks across Canada, and first isolation of the Lyme disease spirochete, *Borrelia burgdorferi*, from the avian tick, Ixodes auritulus. *J Parasitol, 91*(4), 780-790.

Nelson, D. R., Rooney, S., Miller, N. J., & Mather, T. N. (2000). Complement-mediated killing of *Borrelia burgdorferi* by nonimmune sera from sika deer. *Journal of Parasitology, 86*(6), 1232-1238.

Norte, A. C., Ramos, J. A., Gern, L., Núncio, M. S., & Lopes de Carvalho, I. (2013). Birds as reservoirs for *Borrelia burgdorferi* sl in Western Europe: circulation of *B. turdi* and other genospecies in bird–tick cycles in Portugal. *Environmental microbiology, 15(2)*, 386-397.

Nupp, T. E., & Swihart, R. K. (2000). Landscape-level correlates of small-mammal assemblages in forest fragments of farmland. *Journal of Mammalogy, 81*(2), 512-526.

Ogden NH. (2015). Vector borne disease: Box 6-2. Climate change and Lyme disease. pp. 141-143. In: Levy B and Patz J. Climate Change and Public Health. New York: Oxford University Press; 2015. p. 368.

Ogden, N.H., Barker, I.K., Beauchamp, G., Brazeau, S., Charron, D. Maarouf, A., Morshed, M.G., O'Callaghan, C.J., Thompson, R.A., Waltner-Toews, D., Waltner-Toews, M. & Lindsay, L.R. (2006). Investigation of ground level and remote-sensed data for habitat classification and prediction of survival of *Ixodes scapularis* ticks in habitats of southeastern Canada. *Journal of Medical Entomology. 43,* 403-414.

Ogden, N. H., Bigras-Poulin, M., O'Callaghan C, J., Barker, I. K., Kurtenbach, K., Lindsay, L. R., & Charron, D. F. (2007). Vector seasonality, host infection dynamics and fitness of pathogens transmitted by the tick Ixodes scapularis. *Parasitology, 134*(Pt 2), 209-227.

Ogden, N. H., Feil, E. J., Leighton, P. A., Lindsay, L. R., Margos, G., Mechai, S., Michel P., Moriarty, T. J. (2015a). Evolutionary aspects of emerging Lyme disease in Canada. *Appl Environ Microbiol, 81*(21), 7350-7359.

Ogden, N. H., Lindsay, L. R., Hanincova, K., Barker, I. K., Bigras-Poulin, M., Charron, D. F., Thompson, R. A. (2008). 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*(6), 1780-1790.

Ogden, N. H., Lindsay, L. R., & Leighton, P. A. (2013a). Predicting the rate of invasion of the agent of Lyme disease *Borrelia burgdorferi*. *Journal of Applied Ecology, 50*(2), 510-518.

Ogden, N. H., Lindsay, L. R., Morshed, M., Sockett, P. N., & Artsob, H. (2009). The emergence of Lyme disease in Canada. *CMAJ, 180*(12), 1221-1224.

Ogden, N. H., Margos, G., Aanensen, D. M., Drebot, M. A., Feil, E. J., Hanincova, K., Schwartz, I., Tyler, S., Lindsay, L. R. (2011). Investigation of genotypes of *Borrelia*

burgdorferi in *Ixodes scapularis* ticks collected during surveillance in Canada. *Appl Environ Microbiol, 77*(10), 3244-3254.

Ogden, N. H., Mechai, S., & Margos, G. (2013b). Changing geographic ranges of ticks and tick-borne pathogens: drivers, mechanisms and consequences for pathogen diversity. *Front Cell Infect Microbiol, 3*, 46.

Ogden N.H., Koffi, J.K., Lindsay, L.R., Fleming, S., Mombourquette, D.C., Sanford, C., Badcock, J., Gad, R.R., Jain-Sheehan, N., Moore, S., Russell, C., Hobbs, L., Baydack, R., Graham-Derham, S., Lachance, L., Simmonds, K., Scott, A.N. (2015b). Surveillance for Lyme disease in Canada, 2009-2012. *Canada Communicable Disease Report. 41*, 132-145.

Ogden, N. H., & Tsao, J. I. (2009). Biodiversity and Lyme disease: dilution or amplification?. *Epidemics, 1*(3), 196-206.

Ostfeld, R. S., & Keesing, F. (2000). Biodiversity and disease risk: the case of Lyme disease. *Conservation Biology, 14*(3), 722-728.

Pal, U., Li, X., Wang, T., Montgomery, R. R., Ramamoorthi, N., Desilva, A. M., Bao, F., Yang, X., Pypaert, M., Pradhan, D., Kantor, F. S., Telford, S., Anderson, J. F., Fikrig, E. (2004). TROSPA, an *Ixodes scapularis* receptor for *Borrelia burgdorferi*. *Cell,* 119(4), 457-468.

Pan, W., & Wall, M. M. (2002). Small sample adjustments in using the sandwich variance estimator in generalized estimating equations. *Stat Med, 21*(10), 1429-1441.

Patrican, L. A. (1997). Absence of Lyme disease spirochetes in larval progeny of naturally infected *Ixodes scapularis* (Acari: Ixodidae) fed on dogs. *Journal of Medical Entomology, 34*(1), 52-55.

Patton, T. G., Brandt, K. S., & Gilmore, R. D., Jr. (2012). *Borrelia burgdorferi* visualized in *Ixodes scapularis* tick excrement by immunofluorescence. Vector Borne Zoonotic Dis, 12(11), 1000-1003.

Postic, D., Garnier, M., & Baranton, G. (2007). Multilocus sequence analysis of atypical *Borrelia burgdorferi* sensu lato isolates-description of *Borrelia californiensis* sp. nov., and genomospecies 1 and 2. *Int J Med Microbiol, 297*(4), 263-271.

Pulzova, L., & Bhide, M. (2014). Outer surface proteins of Borrelia: peerless immune evasion tools. *Current Protein and Peptide Science, 15*(1), 75-88.

Qiu, W. G., Bruno, J. F., McCaig, W. D., Xu, Y., Livey, I., Schriefer, M. E., & Luft, B. J. (2008). Wide distribution of a high-virulence *Borrelia burgdorferi* clone in Europe and North America. Emerg Infect Dis, 14(7), 1097-1104.

Qiu, W. G., Dykhuizen, D. E., Acosta, M. S., & Luft, B. J. (2002). Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorfe*ri) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. *Genetics, 160*(3), 833-849.

Qiu W. G., Schutzer, S. E., Bruno, J. F., Attie, O., Xu, Y., Dunn, J. J., Fraser, C.M., Casjens, S. R., Luft & B. J. (2004). Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci U S A*, 101, 14150–14155.

Qiu, W. G., & Martin, C. L. (2014). Evolutionary genomics of *borrelia burgdorferi* sensu lato: findings, hypotheses, and the rise of hybrids. *Infect Genet Evol, 27C*.

Radolf, J. D., & Samuels, D. S. (Eds.). (2010). Borrelia: molecular biology, host interaction and pathogenesis. *Horizon Scientific Press*.

Randolph, S.E., and A.D.M. Dobson. 2012. Pangloss revisited: a critique of the dilution effect and the biodiversity-buffers-disease paradigm. *Parasitology 139*, 847–63.

Ras MN, Postic D, Foretz M & Baranton G. (1997). *Borrelia burgdorferi* sensu stricto, a bacterial species "made in the U.S.A."? *Int J Syst Bacteriol,* 47(4), 1112-1117.

Ronquist, F., Huelsenbeck, J., & Teslenko, M. (2011). Draft MrBayes version 3.2 manual: tutorials and model summaries. *Distributed with the software from http://brahms. biology. rochester. edu/software. html*.

Rowe, K. C., Heske, E. J., Brown, P. W., & Paige, K. N. (2004). Surviving the ice: Northern refugia and postglacial colonization. *Proc Natl Acad Sci U S A, 101*(28), 10355-10359.

Rowe, K. C., Heske, E. J., & Paige, K. N. (2006). Comparative phylogeography of eastern chipmunks and white-footed mice in relation to the individualistic nature of species. *Molecular Ecology, 15*(13), 4003-4020.

Rudenko, N., Golovchenko, M., Grubhoffer, L., & Oliver, J. H., Jr. (2011). Updates on *Borrelia burgdorferi* sensu lato complex with respect to public health. *Ticks Tick Borne Dis, 2*(3), 123-128.

Rudenko, N., Golovchenko, M., Lin, T., Gao, L., Grubhoffer, L., & Oliver, J. H., Jr. (2009). Delineation of a new species of the *Borrelia burgdorferi* Sensu Lato Complex, Borrelia americana sp. nov. *J Clin Microbiol, 47*(12), 3875-3880.

Rudnick, D., Ryan, S. J., Beier, P., Cushman, S. A., Dieffenbach, F., Epps, C., Kintsch, J. (2012). The role of landscape connectivity in planning and implementing conservation and restoration priorities. Issues in Ecology.

Rudenko N, Golovchenko M, Hönig V, Mallátová N, Krbková L, Mikulásek P, Fedorova N, Belfiore NM, Grubhoffer L, Lane RS and Oliver JH Jr. (2013). Detection of *Borrelia burgdorferi* sensu stricto ospC alleles associated with human lyme borreliosis worldwide in non-human-biting tick Ixodes affinis and rodent hosts in Southeastern United States. Appl. *Environ. Microbiol*, *79*(5): 1444-1453.

Ružić-Sabljić, E., & Cerar, T. (2017). Progress in the molecular diagnosis of Lyme disease. *Expert review of molecular diagnostics, 17*(1), 19-30.

Saura, S. (2010). Measuring connectivity in habitat mosaics: the equivalence of two existing network indices and progress beyond them. *Community Ecology, 11*(2), 217-222.

Schotthoefer, A. M., & Frost, H. M. (2015). Ecology and Epidemiology of Lyme Borreliosis. *Clinics in laboratory medicine, 35*(4), 723-743.

Schug, M. D., Vessey, S. H., & Korytko, A. I. (1991). Longevity and survival in a population of white-footed mice (Peromyscus leucopus). *Journal of Mammalogy, 72*(2), 360-366.

Schutzer, S. E., Fraser-Liggett, C. M., Casjens, S. R., Qiu, W. G., Dunn, J. J., Mongodin, E. F., & Luft, B. J. (2011). Whole-genome sequences of thirteen isolates of *Borrelia burgdorferi*. *J Bacteriol, 193*(4), 1018-1020.

Schwan, T. G., Piesman, J., Golde, W. T., Dolan, M. C., & Rosa, P. A. (1995). Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *Proceedings of the National Academy of Sciences, 92(7),* 2909-2913.

Seemanapalli, S. V., Xu, Q., McShan, K., & Liang, F. T. (2010). Outer surface protein C is a dissemination-facilitating factor of *Borrelia burgdorferi* during mammalian infection. *PLoS One, 5*(12), e15830.

Seifert, S. N., Khatchikian, C. E., Zhou, W., & Brisson, D. (2015). Evolution and population genomics of the Lyme borreliosis pathogen, *Borrelia burgdorferi*. *Trends in Genetics, 31*(4), 201-207.

Seinost, G., Golde, W. T., Berger, B. W., Dunn, J. J., Qiu, D., Dunkin, D. S., Dykhuizen, D. E., Luft, B. J., Dattwyler, R. J. (1999). Infection with multiple strains of *Borrelia burgdorferi* sensu stricto in patients with Lyme disease. Arch Dermatol, 135(11), 1329-1333.

Simon, J. A., Marrotte, R. R., Desrosiers, N., Fiset, J., Gaitan, J., Gonzalez, A., Millien, V. (2014). Climate change and habitat fragmentation drive the occurrence of *Borrelia burgdorferi*, the agent of Lyme disease, at the northeastern limit of its distribution. *Evol Appl, 7*(7), 750-764.

Singh, S. K., Morbach, H., Nanki, T., Faber, C., Baar, V., & Girschick, H. J. (2004). Differential expression of matrix metalloproteinases and cyclooxygenases in synovial cells exposed to *Borrelia burgdorferi*. *Inflamm Res, 53(12),* 689-696.

Skare, J. T., Shaw, D. K., Trzeciakowski, J. P., & Hyde, J. A. (2016). In vivo imaging demonstrates that *Borrelia burgdorferi* ospC is uniquely expressed temporally and spatially throughout experimental infection. *PLoS One, 11*(9), e0162501.

Smith, R. P., Schoen, R. T., Rahn, D. W., Sikand, V. K., Nowakowski, J., Parenti, D. L., Dykhuizen, D. E., Luft, B. J., Steere, A. C. (2002). Clinical characteristics and treatment outcome of early Lyme disease in patients with microbiologically confirmed erythema migrans. *Annals of internal medicine, 136*(6), 421-428.

Sonenshine, D. E. (2005). The biology of tick vectors of human disease. In Tick-borne diseases of humans (pp. 12-36). *American Society of Microbiology*.

Steere, A. C., Coburn, J., & Glickstein, L. (2004). The emergence of Lyme disease. *J Clin Invest, 113*(8), 1093-1101.

Steere, A. C., Sikand, V. K., Schoen, R. T., & Nowakowski, J. (2003). Asymptomatic infection with Borrelia burgdorferi. *Clin Infect Dis, 37*(4), 528-532.

Štirn, J. (1982). *Manuel des methodes de recherche sur l'environnement aquatique: huitième partie - evulation des modifications des écosystèmes marins dues à la pollution : (directives destinées au projet commun coordonné FAO(CGMP)/PNUE sur la pollution en Méditerranée)*: FAO.

Stewart, P. E., Wang, X., Bueschel, D. M., Clifton, D. R., Grimm, D., Tilly, K., Carroll, J. A.,Weis, J. J., Rosa, P. A. (2006). Delineating the requirement for the *Borrelia*

burgdorferi virulence factor OspC in the mammalian host. *Infect Immun, 74(6),* 3547-3553.

Svardal, H., Rueffler, C., & Hermisson, J. (2011). Comparing environmental and genetic variance as adaptive response to fluctuating selection. *Evolution, 65(9)*, 2492-2513.

Taylor, S., Kumar, L., & Reid, N. (2011). Accuracy comparison of Quickbird, Landsat TM and SPOT 5 imagery for Lantana camara mapping. *Journal of Spatial Science, 56*(2), 241-252.

Telford, S. R., Mather, T. N., Moore, S. I., Wilson, M. L., & Spielman, A. (1988). Incompetence of deer as reservoirs of the Lyme disease spirochete. *The American journal of tropical medicine and hygiene, 39(1),* 105-109.

Tessier, N., Noël, S., & Lapointe, F. J. (2004). A new method to discriminate the deer mouse (Peromyscus maniculatus) from the white-footed mouse (Peromyscus leucopus) using species-specific primers in multiplex PCR. *Canadian Journal of Zoology, 82*(11), 1832-1835.

Thomassen, H. A., Cheviron, Z. A., Freedman, A. H., Harrigan, R. J., Wayne, R. K., & Smith, T. B. (2010). Spatial modelling and landscape-level approaches for visualizing intra-specific variation. *Molecular Ecology, 19*(17), 3532-3548.

Tilly, K., Krum, J. G., Bestor, A., Jewett, M. W., Grimm, D., Bueschel, D., Byram, R., Dorward, D., Vanraden, M. J., Stewart, P., Stewart, P. & Rosa, P. (2006). *Borrelia burgdorferi* OspC protein required exclusively in a crucial early stage of mammalian infection. *Infect Immun, 74*(6), 3554-3564.

Tryon, C., & Snyder, D. (1973). Biology of the eastern chipmunk, *Tamias striatus*: life tables, age distributions, and trends in population numbers. *Journal of Mammalogy, 54*(1), 145-168.

Tuite, A. R., Greer, A. L., & Fisman, D. N. (2013). Effect of latitude on the rate of change in incidence of Lyme disease in the United States. *CMAJ open, 1*(1), E43.

Urwin, R, & Maiden MC. (2003). Multi-locus sequence typing: a tool for global epidemiology. *Trends Microbiol, 11*, 479–487.

Vail, S. G., & Smith, G. (2002). Vertical movement and posture of blacklegged tick (Acari: Ixodidae) nymphs as a function of temperature and relative humidity in laboratory experiments. *Journal of Medical Entomology, 39*(6), 842-846.

Valsangiacomo C, Balmelli T, Piffaretti JC. (1997). A phylogenetic analysis of Borrelia burgdorferi sensu lato based on sequence information from the hbb gene, coding for a histone-like protein. *Int J Syst Bacteriol*, *47*, 1–10.

Van der Linde, M. (1991). Lyme carditis: clinical characteristics of 105 cases. *Scand J Infect Dis Suppl, 77*, 81-84.

Vanwambeke, S. O., Sumilo, D., Bormane, A., Lambin, E. F., & Randolph, S. E. (2010). Landscape predictors of tick-borne encephalitis in Latvia: land cover, land use, and land ownership. *Vector Borne Zoonotic Dis, 10*.

Vollmer, S. A., Bormane, A., Dinnis, R. E., Seelig, F., Dobson, A. D., Aanensen, D. M., James, M.C., Donaghy, M., Randolph, S. E., Feil, E. J., Kurtenbach, K., Margos, G. (2011). Host migration impacts on the phylogeography of Lyme Borreliosis spirochaete species in Europe. *Env Microbiol, 13*.

Vollmer, S. A., Feil, E. J., Chu, C. Y., Raper, S. L., Cao, W. C., Kurtenbach, K., & Margos, G. (2013). Spatial spread and demographic expansion of Lyme borreliosis spirochaetes in Eurasia. *Infect Genet Evol, 14*, 147-155. doi:10.1016/j.meegid.2012.11.014

Walls, J. J., Greig, B., Neitzel, D. F., & Dumler, J. S. (1997). Natural infection of small mammal species in Minnesota with the agent of human granulocytic ehrlichiosis. *J Clin Microbiol, 35*(4), 853-855.

Walter, K. S., Carpi, G., Evans, B. R., Caccone, A., & Diuk-Wasser, M. A. (2016). Vectors as Epidemiological Sentinels: Patterns of Within-Tick *Borrelia burgdorferi* Diversity. *PLoS Pathog, 12*(7), e1005759. doi:10.1371/journal.ppat.1005759.

Wang G, Liveris D, Mukherjee P, Jungnick S, Margos G and Schwartz I. (2014). Molecular Typing of *Borrelia burgdorferi*. Curr Protoc Microbiol. 34: 12C.5.1–12C.5.31.

Wang G, van Dam AP, Dankert J. (2000). Two distinct ospA genes among Borrelia valaisiana strains. *Res Microbiol*, 151:325–331.

Wang, N., Dykhuizen, D. E., Qiu, W., Dunn, J. J., Bosler, E. M., & Luft, B. J. (1999). Genetic diversity of ospC in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics, 151*(1), 15-30.

Wang, X., Huang, Y., Niu, S. B., Jiang, B. G., Jia, N., van der Geest, L., Ni, X. B., Sun, Y., Cao, W. C. (2014). Genetic Diversity of Salp15 in the *Ixodes ricinus* Complex (Acari: Ixodidae). *PLoS One, 9(4*), *e94131.*

Waters, J. H. (1963). Biochemical Relationships of the Mouse *Peromyscus* in New England. *Systematic Zoology, 12*(3), 122-133.

Wójcik-Fatla, A., Zając, V., Sawczyn, A., Sroka, J., Cisak, E., & Dutkiewicz, J. (2016). Infections and mixed infections with the selected species of Borrelia burgdorferi sensu lato complex in Ixodes ricinus ticks collected in eastern Poland: a significant increase in the course of 5 years. *Experimental and Applied Acarology, 68*(2), 197-212.

Wood, C. L., & Lafferty, K. D. (2013). Biodiversity and disease: a synthesis of ecological perspectives on Lyme disease transmission. *Trends in Ecology & Evolution, 28*(4), 239-247.

World Health Organization. (2000). *WHO report on global surveillance of epidemic-prone infectious diseases*. Retrieved from.

Wormser, G. P. (2006). Clinical practice. Early Lyme disease. N Engl J Med, 354(26), 2794-2801.

Wormser, G. P., Brisson, D., Liveris, D., Hanincova, K., Sandigursky, S., Nowakowski, J., Nadelman, R. B., Ludin, S., Schwartz, I. (2008). *Borrelia burgdorferi* genotype predicts the capacity for hematogenous dissemination during early Lyme disease. J Infect Dis, 198(9), 1358-1364.

Wormser, G. P., Kaslow, R., Tang, J., Wade, K., Liveris, D., Schwartz, I., & Klempner, M. (2005). Association between human leukocyte antigen class II alleles and genotype of Borrelia burgdorferi in patients with early lyme disease. *J Infect Dis, 192*(11), 2020-2026.

Xu, Q., McShan, K., & Liang, F.T. (2007) Identification of an ospC operator critical for immune evasion of *Borrelia burgdorferi*. *Mol Microbiol 64: 220–231*.

Zhang, L., Zhang, Y., Adusumilli, S., Liu, L., Narasimhan, S., Dai, J., Zhao, Y. O. & Fikrig, E. (2011). Molecular interactions that enable movement of the Lyme disease agent from the tick gut into the hemolymph. *PLoS Pathog, 7(6), e1002079.*

Zeller, K. A., McGarigal, K., & Whiteley, A. R. (2012). Estimating landscape resistance to movement: a review. *Landscape Ecology, 27*(6), 777-797.

Zinsstag, J., Schelling, E., Waltner-Toews, D., & Tanner, M. (2011). From "one medicine" to "one health" and systemic approaches to health and well-being. *Prev Vet Med, 101*(3), 148-156.