
Link to official URL (if available):
http://dx.doi.org/10.1016/j.meegid.2011.07.022
Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato

Gabriele Margos¹*, Stephanie A. Vollmer¹, Nicholas H. Ogden², Durland Fish³

¹Department of Biology and Biochemistry, University of Bath, Claverton Down, Bath BA2 7AY, UK

²Zoonoses Division, Centre for Food-borne, Environmental and Zoonotic Infectious Diseases, Public Health Agency of Canada, Ottawa, Canada

³Department of Epidemiology and Public Health, Yale School of Medicine, Yale University, New Haven, CT 06520

*corresponding author:

Department of Biology and Biochemistry

3 South

University of Bath

Claverton Down

Bath BA2 7AY

U.K.

Telefon: +44-1225-385116

Fax: +44-1225-38

e-mail gm250@bath.ac.uk
Abstract

To understand the population structure and dynamics of bacterial microorganisms, typing systems that accurately reflect the phylogenetic and evolutionary relationship of the agents are required. Over the past 15 years multilocus sequence typing schemes have replaced single locus approaches, giving novel insights into phylogenetic and evolutionary relationships of many bacterial species and facilitating taxonomy. Since 2004, several schemes using multiple loci have been developed to better understand the taxonomy, phylogeny and evolution of Lyme borreliosis spirochetes and in this paper we have reviewed and summarized the progress that has been made for this important group of vector-borne zoonotic bacteria.

KEYWORDS: BORRELLIA BURGDORFERI, EVOLUTION, PHYLOGENY, MOLECULAR ECOLOGY, IXODES, TICKS, MULTILOCUS SEQUENCE TYPING, MLST

ABBREVIATIONS:

IGS – INTERGENIC SPACER
LB – LYME BORRLLIOSIS
MLST/MLSA – MULTILOCUS SEQUENCE TYPING/MULTILOCUS SEQUENCE ANALYSIS
MW – MIDWEST
NE – NORTHEAST
OSP – OUTER SURFACE PROTEIN
SLV/DLV/TLV – SINGLE LOCUS VARIANT/DUOUBLE LOCUS VARIANT/TRIPLE LOCUS VARIANT
SNP – SINGLE NUCLEOTIDE POLYMORPHISM
ST – SEQUENCE TYPE
1. INTRODUCTION

Tick-borne diseases are of increasing public health concern because of range expansions of both vectors and pathogens (Daniel et al., 2003; Falco et al., 1995; Ogden et al., 2008b). To understand these processes and to predict future trajectories, detailed data on the contemporary population structure and on the evolutionary and demographic histories that have shaped the populations are essential. Population structure, evolutionary and demographic processes of microbial pathogens may best be inferred using genetic data with neutral variation. Such data together with information on host associations are critical to understand the dynamics of tick-borne disease agents and to form hypothesis concerning past and future spread.

Lyme borreliosis (LB) is the most prevalent vector-borne disease in the Holartic region (Dennis and Hayes, 2002). Due to the pattern and breadth of the ecological niches occupied by its members, the LB group of spirochetes constitutes an ideal system to investigate the contributions of host and vectors in pathogen demographic processes. In addition, major advances in sequencing technologies and the development of sophisticated typing tools for bacterial pathogens have greatly enhanced the potential to infer robust phylogenies and to deduce more accurately the evolutionary relationships of micro-organisms. In particular, targeted gene amplification and sequence analysis of several housekeeping genes, termed multilocus sequence typing or multilocus sequence analysis (MLST/MLSA) and, more recently, genome-wide detection of single nucleotide polymorphisms (SNPs) have made major contributions to advancing knowledge in bacterial population genetics, phylogenetics and molecular taxonomy (Aanensen and Spratt, 2005; Bishop et al., 2009; Hall, 2007; Harris et al., 2010; Holt et al., 2008; Maiden, 2006). In this review we are focussing
on progress that has been made in recent years using molecular methods including MLST and MLSA to study population genetics, molecular taxonomy, phylogenetics and the evolution of the LB group of spirochetes (also referred to as *Borrelia burgdorferi* sensu lato (s.l.) species complex). Although we acknowledge that not all species belonging to this species complex cause LB, we prefer to use the term ‘LB group of spirochetes’ (instead of *B. burgdorferi* s.l.) to refer to the whole group as this simplifies distinguishing *B. burgdorferi* s.l. from *B. burgdorferi* sensu stricto (the species to which we will refer hereafter as *B. burgdorferi*).

The species complex currently consists of 18 proposed and confirmed species (Margos et al., 2010; Rudenko et al., 2009a; Rudenko et al., 2009b) (Table 1), several of which can cause LB in humans (or Lyme disease). LB species vary in their geographic distribution, host specificity and ability to cause disease in humans. Clinically the different pathogenic *Borrelia* spp. are of interest as they have been associated with different disease symptoms which may be observed in the late stages of the condition. For example, *B. afzelii* is most frequently linked with skin manifestations, *B. garinii* and *B. bavariensis* with neuroborreliosis, and *B. burgdorferi* with arthritic symptoms (Canica et al., 1993; Ornstein et al., 2001; Randolph, 2008; Rijpkema et al., 1997; Stanek and Strle, 2009; Steere et al., 1986; van Dam, 2002). Some species, such as *B. lusitaniae*, have only occasionally been associated with human disease while for others, such as *B. valaisiana*, the status is uncertain because they have high regional prevalence in Europe but have rarely been isolated from humans (Collares-Pereira et al., 2004; Diza et al., 2004). It has been suggested that not all strains/genotypes within a species cause disseminated disease (Baranton et al., 2001; Seinost et al., 1999; Wilske et al., 1996; Wilske et al., 1993; Wormser et al.,
and it is, therefore, of epidemiological and clinical relevance to identify the geographic range of LB species and the spatial distributions of their genotypes.

2. Ecology of LB group of spirochetes

Due to the obligate parasitic lifestyle of LB spirochetes, their biology is intimately linked to that of their invertebrate and vertebrate hosts which also broadly defines their ecological niches (Kurtenbach et al., 2002b). The ecological niche diversity of different species varies in the degree of specialization (from generalist to specialised strategies) in terms of host and vector adaptation and this influences the geographic distribution at species and population levels. There are several excellent recent reviews regarding the ecology of LB spirochetes, describing in detail host and vector interactions (Gern, 2008; Gern and Humair, 2002; Kurtenbach et al., 2006; Masuzawa, 2004; Piesman and Gern, 2004; Tsao, 2009). Here we will only briefly describe the general ecology of LB spirochetes.

The life cycle of the LB group of spirochetes is a dynamic interplay between bacteria, reservoir hosts and vectors which is confounded by landscape and climatic factors impacting host and vector ecology (Figure 1, (Kurtenbach et al., 2006)). All known vectors of LB spirochetes belong to the genus *Ixodes* and these ticks are three host ticks, i.e. they have three feeding stages (larvae, nymphs and adult females) each utilizing a different host, although not necessarily a different host species. Except in the case of nidicolous (nest-living) tick vectors, adult female ticks prefer large animals, such as deer, as hosts which are considered not susceptible to *Borrelia* infection (Telford et al., 1988). The preference of both immature stages for small to medium sized vertebrates (mammals, birds or lizards) is essential for maintaining the bacteria in its natural transmission cycles. The bacteria are taken up during a
bloodmeal from an infected and infectious host, are maintained transstadially during
the moulting process and are then transmitted to other hosts during the subsequent
bloodmeal during the next life stage (Gern and Humair, 2002). Other means of
transmission are co-feeding transmission (between neighbouring ticks feeding on a
susceptible or non-susceptible host) (Ogden et al., 1997) and transovarial transmission
((Gern and Humair, 2002) and references therein), although the latter may depend on
the tick species as it has not been experimentally demonstrated for *I. scapularis* and *I.
persulcatus* (Nefedova et al., 2004; Patrican, 1997). However, relapsing-fever like
spirochetes (e.g. *B. miyamotoi*) are transmitted transovarially in *Ixodes* ticks and occur
sympatrically with LB group spirochetes, which may explain some or perhaps all
observations of transovarial transmission (Piesman, 2002; Scoles et al., 2001).

The main vectors transmitting LB spirochetes to humans are members of the
*Ixodes persulcatus* species complex and are generalist feeders (i.e. they have a wide
host range) that follow an ambush strategy for host seeking and are widely distributed
in the environment (Balashov, 1972; Loye and Lane, 1988; Xu et al., 2003). These are
*I. ricinus* in Europe, *I. persulcatus* in Eastern Europe and Asia, *I. scapularis* and *I.
pacificus* in North America. Other experimentally confirmed vector-competent *Ixodes*
species are nidicolous to varying degrees, i.e. they reside in the burrows of their hosts,
and, having a more restricted host preference, rarely bite humans (see review by
(Eisen and Lane, 2002)). This raises the question as to whether the LB spirochetes
transmitted by nidicolous vectors are non-pathogenic for humans, or whether they are
pathogenic but rarely cause disease because the ticks that transmit them rarely
encounter humans.

More than 100 vertebrate species have been identified that can act as reservoir
hosts for LB spirochetes including rodents (wood mice, wood rats, voles, dormice,
squirrels, chipmunks, rats), insectivores (shrews, hedgehogs), raccoons and several
bird species (Masuzawa, 2004; Piesman and Gern, 2004). For other species such as
foxes or badgers only limited information is available and it is uncertain whether they
constitute reservoir hosts (Gern and Sell, 2009; Matuschka et al., 2000; Miyamoto and
Masuzawa, 2002), although domestic dogs have been reported to be reservoir
competent (Mather et al., 1994). Not all vertebrate hosts are permissive or equally
efficient as reservoir hosts for all *Borrelia* species. The basic reproduction number $R_0$
serves as a measure of fitness of different LB group species in different host-tick
communities (reviewed by (Randolph, 1998) and (Tsao, 2009)). For some LB group
species only certain host species are able to support completion of the entire
transmission cycle (from a vector tick through the host to the next vector tick)
(Kurtenbach et al., 2002a). In Europe, these ‘host associations’ have been well studied
and they are an important component of the ecology of LB spirochete species. Most
of the LB group species in Europe are transmitted by the generalist tick, *I. ricinus*,
which feeds on birds as well as on rodents or other medium sized mammals, rendering
the tick a ‘mixing vessel’ for different strains and species. Therefore, the host
associations described in Europe are not driven by adaptation to an endophilic tick
with a narrow host preference but are truly host driven (Humair and Gern, 2000;
Kurtenbach et al., 1998b). Several lines of evidence support the notion of host
association: 1) Experimental evidence has shown that these host associations match
the ability of LB group species to deflect complement mediated lysis of the
corresponding reservoir hosts (Kurtenbach et al., 2002b; Kurtenbach et al., 1998a;
Lane and Quistad, 1998; Ullmann et al., 2003). 2) *B. afzelii* and *B. bavariensis* have
been shown to be transmitted through rodents while *B. garinii* and *B. valaisiana* are
transmitted through avian reservoir hosts (Dubska et al., 2009; Hanincova et al.,
This does not mean that *B. garinii* infections cannot be found in mice, as bird adapted outer surface protein A (OspA) serotype 6 strains have been found in internal organs of *Apodemus* mice, but these strains are not transmitted to vector competent ticks feeding on such infected mice and, therefore, represent dead-end hosts (Kurtenbach et al., 1998a; Kurtenbach et al., 2002a). Mechanisms permitting the transmission to hosts of such host complement incompatible LB spirochete species have been suggested (Kurtenbach et al., 2002a). Recent evidence supports the view that host associations substantially shape *Borrelia* populations by impacting their migration patterns and geographical distributions (Kurtenbach et al., 2006; Vollmer et al., 2011). The compatibility of spirochetes with tick vectors has not been studied in so much detail. While many *Ixodes* tick species are able to transmit several species of *Borrelia* (Table 1), it seems that certain *Borrelia*-vector associations are not compatible or less efficient (e.g. (Dolan et al., 1998; Masuzawa et al., 2005)). Thus, vector competence - or the lack thereof - has implications on the geographic distribution of these species (see geographic distribution). Consequently, *Borrelia* populations are shaped by the dynamics and demographic processes of host and vector populations, host and vector immune responses and extrinsic abiotic factors (e.g. temperature, climate, landscape connectivity) affecting host and vector populations and contact between them which together determine $R_0$ for each species and strain of the bacterium (Figure 1). Diversity in *Borrelia* populations arises by mutation, recombination, drift and natural selection. It has been suggested that mutation rates are low as *Borrelia* are very slow growing bacteria (Hoen et al., 2009). Genetic drift may also predominate when...
effective population sizes, $N_e$, are small (as has been suggested for $B. burgdorferi$ (Qiu et al., 2002)), this may weaken natural selection and introduce stochastic effects into allele frequencies of populations (Page and Holmes, 1998). The signature of all these processes can be inferred from genetic information obtained from present day samples but as different processes can lead to similar effects, caution needs to be exercised when interpreting data (Frank, 2002).

3. Typing tools for the LB group of spirochetes

When $B. burgdorferi$ was discovered and described, it was assumed to be a single species (Burgdorfer et al., 1982; Johnson et al., 1984). The use of genome fingerprinting and other methods soon showed that the bacteria were highly diverse and in fact represented a species complex (Liveris et al., 1995; Marconi and Garon, 1992; Mathiesen et al., 1997; Postic et al., 1994; Wilske et al., 1991). Phenotypic typing tools were developed to reveal intraspecific diversity which included serotyping or multilocus enzyme electrophoresis (MLEE) (Boerlin et al., 1992; Wilske et al., 1991; Wilske et al., 1996; Wilske et al., 1995; Wilske et al., 1993). This topic is reviewed excellently by Wang and co-authors (Wang et al., 1999b) and here we concentrate on single and multilocus sequence analyses.

Sequences of single gene loci have been popular for ecological, population, epidemiological and evolutionary studies of the LB group of spirochetes. Many different genes and loci have been targeted in studies depending on the level of variation and the discriminatory power required and which species were being investigated. These included intergenic spacer (IGS) regions, the $rrs$ (16S rRNA) locus, the plasmid located genes encoding the outer surface proteins A and C ($ospA$, $ospC$), decorin-binding protein A ($dbpA$), the chromosomally located housekeeping
genes recombinase A (recA), groEL, hbb or flagellin B (flaB) (Casati et al., 2004; Dykhuizen and Baranton, 2001; Fukunaga et al., 1996c; Liveris et al., 1995; Marconi et al., 1995; Michel et al., 2004; Park et al., 2004; Postic et al., 1994; Schulte-Spechtel et al., 2006; Valsangiacomo et al., 1997; Will et al., 1995; Wilske et al., 1996).

3.1 Interspecies Studies

For species definition and evolutionary studies, conserved loci or intergenic spacer have been employed. flaB has been popular for evolutionary studies and species identification because the flaB gene is present in relapsing fever spirochetes, which can be used as an outgroup to root phylogenetic trees. This conserved locus was used to create an early and reasonably complete evolutionary tree of the LB group of spirochetes (Fukunaga et al., 1996c). Some groups now use this and other conserved loci (e.g. 23S, hbb) to screen field-collected questing ticks (by real-time or conventional PCR) and to establish infection prevalences with LB species, as well as with relapsing-fever like spirochaete species such as B. miyamotoi which infects hard bodied ixodid ticks worldwide (Barbour et al., 2010; Fukunaga et al., 1995; Herrmann and Gern, 2010; Ogden et al., 2011; Portnoi et al., 2006).

The region encoding the ribosomal RNAs (rRNA) has been popular in studies of LB spirochaetes where different regions have been used for various purposes and species. The 16S (rrs) subunit, has been used in evolutionary and speciation studies (e.g. (Fukunaga et al., 1996a; Le Fleche et al., 1997).

Approximately 2 kb downstream of a single copy of the 16S rRNA small subunit are tandemly repeated copies of the 23S-5S (rrl-rrf) large subunits (Schwartz et al., 1992). The IGS between the 5S and 23S (rrf-rrl) of the repeated pairs is approximately 200-250 bp and this organization of rRNA genes appears to be unique.
to the LB group spirochetes (Gazumyan et al., 1994; Schwartz et al., 1992). The 5S-23S spacer is possibly the most common sequence-based method for LB group species identification in Europe and approaches have recently been developed using quantitative PCR to screen questing ticks (Postic et al., 1994; Postic et al., 1998; Strube et al., 2010). Diversity at this locus has also been investigated using reverse line blot, a key method in epidemiological studies of LB species due to it being a rapid and reliable method for detecting and typing mixed infections of different Borrelia species in field-collected tick or host samples. It uses PCR products of the 5S-23S IGS region for hybridization to membrane bound oligonucleotides that are specific for different LB group species. This method was first used to identify the prevalence of different LB group species in ticks in The Netherlands (Rijpkema et al., 1997). Reverse line blot was better suited than some other methods for characterising mixed species infections and partly for this reason it was a key method in identifying the patterns of host specialization (Hanincova et al., 2003b; Kurtenbach et al., 2001; Kurtenbach et al., 1998a). A problem was that this method was unable to distinguish ecotypes of B. garinii (bird or rodent associated which are now considered different species, (Margos et al., 2009)) which may have confused some conclusions of host associations.

3.2 Intraspecies Genotyping

For intraspecies studies loci that provide good level of polymorphism have been widely used, such as the 16S-23S (rrs-rrl) IGS or outer surface protein (osp) encoding loci for B. burgdorferi in North America (Bunikis et al., 2004; Girard et al., 2009; Hamer et al., 2010; Hanincova et al., 2008a; Liveris et al., 1995; Marconi et al., 1995;
Ogden et al., 2008b; Postic et al., 1994). However, these are not necessarily useful for species identification or for intraspecies studies of other LB group species.

Outer surface proteins are variable and have, for this reason, often been used for population studies. *ospA*, located on a 49- to 70-kb linear plasmid, called lp54 in *B. burgdorferi* (Barbour and Garon, 1988), revealed differences in the levels of homogeneity of LB species. It was observed that there is great variation in *ospA* in *B. garinii* while there is much homogeneity in some other species such as *B. burgdorferi* or *B. afzelii* which is consistent with serotyping studies (Wilske et al., 1996). This locus has also been used to reveal rare horizontal gene transfer between species (Rosa et al., 1992; Wang et al., 2000).

*ospC* is located on a 26-kb circular plasmid (Sadziene et al., 1993) and has been described as the locus with the highest degree of variation (Jauris-Heipke et al., 1995; Qiu et al., 2004; Theisen et al., 1993). This locus is rarely used for species determination because, while there may be species specific motifs (Fukunaga and Hamase, 1995; Jauris-Heipke et al., 1995), recombination and plasmid exchange means that strains of the same species do not always cluster monophyletically in phylogenies (Kurtenbach et al., 2002a; Lin et al., 2002; Margos et al., 2009) (Figure 2B). However, due to the high level of variation, *ospC* has been frequently used in population studies within species, most notably within *B. burgdorferi* (Barbour and Travinsky, 2010; Hanincova et al., 2008a; Marti Ras et al., 1997; Qiu et al., 2002), and the study of *ospC* may be useful in identifying ecological traits such as host-species associations (Ogden et al., 2011) as its expression is important for tick-to-host transmission (Piesman and Schwan, 2010).

Population genetics studies on *B. burgdorferi* in the Northeastern (NE) USA have suggested that *ospA* is in linkage disequilibrium with *ospC*, a gene on a different
plasmid, and the 16S-23S IGS (Qiu et al., 1997) while more recent studies have shown that this may be related to the spatial scale of sampling as geographic variation in linkage pattern were found (Hellgren et al., 2011; Travinsky et al., 2010). In addition, horizontal transfer has been demonstrated for many plasmid-encoded loci, whole plasmids and also for genes on the main chromosome although it needs to be emphasized that these are likely to be rare events (Barbour and Travinsky, 2010; Qiu et al., 2004; Vitorino et al., 2008; Wang et al., 1999a) unpublished). While Qiu and co-authors (Qiu et al., 2004), using almost exclusively plasmid-located loci, found a higher rate of recombination than mutation (ratio 3:1), studies using chromosomally located housekeeping genes found higher mutation than recombination rates with an r/m of 1:100 to 1:25, strongly suggesting that the linear chromosome is well suited for studies investigating evolutionary and population relationships of LB spirochetes ((Vitorino et al., 2008), Vollmer et al. unpublished).

Different loci tended to be preferred in North America or Asia compared to Europe. Many studies conducted in the USA, where B. burgdorferi is the only species causing human disease, have focused on ospC, the 16S-23S IGS or a combination of these and additional loci (Brisson and Dykhuizen, 2004; Brisson et al., 2010; Bunikis et al., 2004; Girard et al., 2009; Hanincova et al., 2008a; Liveris et al., 1995; Marti Ras et al., 1997; Ogden et al., 2008b; Qiu et al., 2002; Wang et al., 1999; Wormser et al., 1999). In Europe several Borrelia species are prevalent and all four major disease-causing species (i.e. B. afzelii, B. garinii, B. burgdorferi and B. bavariensis) are endemic in populations of I. persulcatus-group ticks (Gern, 2008). For this reason species definition has been the key for epidemiological and ecological studies, and thus, in Europe ospA and the 5S-23S IGS region have been most commonly used (summarized by (Rauter and Hartung, 2005)). In Asia, species identification was often
the major aim of studies and a variety of loci have been used including loci favoured in Europe as well as more conserved loci, such as flaB and 16S rRNA (Masuzawa, 2004). This is most likely because fewer population genetic studies have been completed and species prevalence is of primary importance over such a large area considering the broad spectrum of species found across the continent.

3.3 Typing schemes using multiple loci

Since 2004 several multilocus schemes have been developed to investigate the phylogenetic relationship of the LB spirochetes. The greater amount of genetic information obtained from several loci permits determination of more subtle differences in and between species.

MLST schemes were originally designed to utilise regions of housekeeping genes that evolved at a moderate speed to capture the intermediate relationship within bacterial species (Figure 3) (Maiden, 2006; Maiden et al., 1998). While this means that the number of polymorphic sites per gene region is usually low, by combining multiple loci the discriminatory power is high. Traditionally, internal fragments of housekeeping genes, approximately 450-500 bp long, were selected and kept in-frame. The genes were chosen throughout the genome to avoid any local bias that may occur in the bacterial genome. Another criterion was that the chosen housekeeping genes should also be flanked by genes known to have similar functions as there may be linkage between adjacent genes. If genes next to the selected housekeeping gene are under strong selection pressures, this may influence the neighbouring genes. Finally, genes should have a similar level of genetic diversity so that each gene provides a similar contribution to phylogenetic analyses and no single
gene dominates a tree generated by use of the concatenated sequences of the selected
housekeeping genes (Urwin and Maiden, 2003).

One central problem when attempting to understand relationships among
bacterial species or populations is posed by genetic recombination. This is because
there is a possibility that a single locus representing a particular strain may have
undergone a recombination event with another strain or species and this locus would
not be representative of the “true” evolutionary pathways of that particular strain
genome. In other words, the use of a single locus will infer the evolution of this
particular locus but not necessarily the evolution of the organism as a whole. MLST
schemes aim at overcoming this problem by combining several, often seven, loci that
are scattered across the genome. Thus, if one region of the genome has undergone
recombination only one or two of the seven genes may be affected. This means
primarily that if recombination is occurring it is easier to identify it by comparing
base pair changes in the loci of closely related strains or the linkage between genes
(Didelot and Falush, 2007; Feil et al., 2000). Secondly, in MLST schemes each allele
of each gene is given a unique number so that isolates can be characterised by a multi-
integer number called an allelic profile. This means that, regardless of whether a
particular strain differs from another strain in a single locus by a single base pair
(indicative of mutation) or many base pairs (indicative of recombination), in terms of
the allelic profile, the strains will only differ by a single integer number. Thus
analysing strains using their allelic profiles will buffer the distorted effect
recombination may have on phylogenetic inferences or any other analyses.

Once the genes have been selected and the MLST scheme is in place,
sequence data, strain information and allelic profiles are compiled by “virtual isolate
collections centres” in the form of online databases (Urwin and Maiden, 2003) such as
Each unique allelic profile is given a unique number called a sequence type (ST) allowing for easy reference to particular isolates. The original aim of the MLST concept was to enhance clinical diagnosis, epidemiological monitoring, and population studies (Urwin and Maiden, 2003) but the MLST concept has since been broadened to include the analysis of closely related species and this approach has been named multi-locus sequence analysis (MLSA) (Gevers et al., 2005; Hanage et al., 2006; Hanage et al., 2005). MLSA was developed with the aim of allowing for rapid and robust hierarchical classification of all prokaryotic species (Gevers et al., 2005) and has been raised as a solution to the time-consuming and complicated method of prokaryote species definition by DNA-DNA hybridization (Bishop et al., 2009; Gevers et al., 2005). Recently a website has been developed to allow the species identification of unknown isolates of Streptococcal species (thought to be a taxonomically challenging group) by entering the sequence data of seven gene fragments (Bishop et al., 2009).

For the LB group of spirochetes five schemes using multiple loci have been developed (Table 2) (Bunikis et al., 2004; Margos et al., 2008; Qiu et al., 2004; Richter et al., 2006; Rudenko et al., 2009a) and recently a mixture of two typing schemes was used (Gomez-Diaz et al., 2011). Three of these schemes have been used as an alternative to DNA-DNA hybridization, i.e. to delineate new species (Chu et al., 2008; Margos et al., 2010; Margos et al., 2009; Postic et al., 2007; Richter et al., 2006; Rudenko et al., 2009a; Rudenko et al., 2009b). Schemes by Bunikis et al. (2004), Qiu et al. (2004) and Rudenko et al. (2009a) have tended to focus on species found in the United States, with two focusing almost entirely on B. burgdorferi (Attie et al., 2007; Brisson et al., 2010; Bunikis et al., 2004; Qiu et al., 2004). However, most of these schemes did not adhere to the strict criteria set out by Urwin and
Maiden (2003), described above, because they combine a variety of gene types including slowly evolving housekeeping genes, non-coding regions, or fast evolving plasmid encoded loci. The loci differ in terms of the selective processes acting upon them, the number of variable sites within these loci as well as the gene category. This may lead to problems when inferring phylogenies as combining sequence data that are heterogeneous, as loci of different functional categories frequently are, can reduce the power of phylogenetic inference algorithms or even produce erroneous phylogenies (Huelsenbeck et al., 1996). Furthermore, the use of the 5S-23S IGS region as well as ospA means there is no species available to act as an outgroup to root a phylogeny and to allow for evolutionary inferences. For the MLSA scheme based on housekeeping genes (Margos et al., 2008), a website (borrelia.mlst.net) is maintained at Imperial College London, UK. It currently contains data for approximately 1,200 Borrelia strains comprising most of the described LB group species which have been resolved into >300 STs from Europe, Asia, and North America. The accumulative nature of MLST databases and the additional information gathered (e.g. geographic coordinates) makes it an attractive instrument to understand intra- and inter-species relationships on a global and regional scale.

4. Borrelia taxonomy

Bacterial taxonomy is a scientific discipline in flux (Gevers et al., 2006). For many years in bacterial systematics the accepted species definition was that a species would include strains with greater than 70 % homology when tested by DNA-DNA hybridization and a ΔTm of 5°C or less. Below the value of 70 % homology strains were considered different species (Wayne et al., 1987). DNA-DNA hybridization requires a specialized laboratory and the number of laboratories that can perform this
analysis worldwide is limited. There are also questions about the interpretation and reproducibility of the method (Stackebrandt and Ebers, 2006). As this method is complicated, sequencing of the 16S rRNA locus and phylogenetic analysis was a valuable and widely used tool for bacterial classification. Both these methods, however, lacked sensitivity at the species level (Staley, 2006). Multilocus sequence analysis (MLSA), the genus-wide application of MLST, was proposed as an alternative to DNA-DNA hybridization and this technique is increasingly used in bacterial classification (Bishop et al., 2009; Gevers et al., 2006).

For LB group spirochetes, in addition to DNA-DNA hybridization and 16S sequences, analyses of the 5S-23S IGS have also served for species and strain typing (Postic et al., 1994). Several species have been defined using these methods including *B. burgdorferi* B31, *B. afzelii* VS461, *B. garinii* 20047, *B. japonica* HO14, *B. valaisiana* VS116 and *B. lusitaniae* PotiB2 (Baranton et al., 1992; Johnson et al., 1984; Kawabata et al., 1993; Le Fleche et al., 1997; Wang et al., 1997). In MLSA analyses these LB species cluster monophyletically at the end of long branches separating the different species (Margos et al., 2009; Richter et al., 2006) (see Figure 2 A).

For *Borrelia* taxonomy, the different schemes using multiple loci employed varying loci (Table 2). These schemes have been used to define several new *Borrelia* species (i.e. *B. spielmanii*, *B. californensis*, *B. carolinensis*, *B. americana*, *B. yangtze*, *B. bavariensis* and *B. kurtenbachii*) by genetic distance analyses (Chu et al., 2008; Margos et al., 2010; Margos et al., 2009; Postic et al., 2007; Richter et al., 2006; Rudenko et al., 2009a; Rudenko et al., 2009b). Richter and colleagues (Richter et al., 2006) and Postic and colleagues (Postic et al., 2007) compared the genetic distances of strains, based on the concatenated sequence of multiple genes, to the corresponding
whole DNA-DNA hybridization genetic distance data and determined a ‘cut-off’
value for species determination for their scheme (Postic et al., 2007; Richter et al.,
2006). To determine this cut-off value, two European *B. burgdorferi* strains, NE49
and Z41293, which were ‘borderline’ *B. burgdorferi* stains in DNA-DNA
hybridization, were used (Postic et al., 2007). A recently proposed 19th species, *B.
finlandensis* (Casjens et al., 2011), belongs to this group of ‘borderline’ *B. burgdorferi*
strains as determined by MLSA (see borrelia.mlst.net). While for the Richter-scheme
the cut-off value was determined to be 0.21, using the same strains, Margos and co-
authors (Margos et al., 2009) determined a cut-off value of 0.170 for the scheme
based on eight chromosomally located housekeeping genes. This scheme permitted *B.
bavariensis*, a rodent-associated ecotype previously named *B. garinii* OspA serotype
4, to be distinguished from other bird-associated *B. garinii* strains. This MLSA
system enabled Takano and co-authors to determine that in Japan most human-
pathogenic *Borrelia* isolates were phylogenetically closer related to the rodent-
adapted sequence types ST84 and ST85 (*B. bavariensis*) than to bird-associated *B.
garinii* (Takano et al., 2011).

5. Geographic distribution

The LB species are not evenly distributed across the globe (Figure 4). Host
specialization and/or vector compatibility of the LB spirochetes are likely to influence
the global distribution of different spirochetal species. In Europe, eight species have
been recorded of which three (*B. garinii*, *B. afzelii*, and *B. bavariensis*) are also found
throughout Asia (Baranton et al., 1992; Korenberg et al., 2002; Masuzawa, 2004;
Takano et al., 2011). *B. valaisiana*, which occurs sympatrically with *B. garinii* in
Europe, has rarely been found in *I. persulcatus* and appears to be absent in Russia and
most of Asia except for a single strain that was found in *I. columnae* in Japan (Bormane et al., 2004; Korenberg et al., 2002; Masuzawa, 2004). Similarly, *B. burgdorferi* has not been found in *I. persulcatus*, a main vector of *B. afzelii, B. garinii* and *B. bavariensis*-like strains in Russia and Asia. Furthermore, NT29 strains of *B. garinii* (which are rodent-adapted and genetically closely related to *B. bavariensis* (unpublished)) occur in Russia and Asia but have not been found in *I. ricinus* (Korenberg et al., 2002; Masuzawa et al., 2005). These authors concluded that the distribution range of NT29 strains is associated with that of a single vector species, *I. persulcatus*. This is interesting in view of the close phylogenetic relationship that has been found for *B. bavariensis* (which is transmitted by *I. ricinus*) and rodent-adapted *B. garinii* from Asia (Takano et al., 2011) and could provide an attractive system to investigate *Ixodes* vector adaptations of *Borrelia* species. Species with a localised distribution are *B. tanukii, B. turdi, and B. japonica* in Japan (Fukunaga et al., 1996b) and *B. lusitaniae* around the Mediterranean Basin. Lizards of the family Lacertidae have been identified as important hosts for the latter species (Amore et al., 2007; Richter and Matuschka, 2006; Younsi et al., 2005). Whether or not other hosts are reservoir competent has not been shown but, occasionally, infections of questing ticks with *B. lusitaniae* have been described in other parts of Europe such as Poland and Latvia (Vollmer et al., 2010; Wedeka and Skotarczak, 2005). *B. garinii* possibly has the broadest distribution of all the LB group spirochetes. Not only is it found in forested regions across Eurasia, it is also maintained in sea bird colonies by the tick vector, *I. uriae*. This means it is also found in many far reaching sites including arctic regions and colonies off the east coast of Canada (Duneau et al., 2008; Smith et al., 2006). At first sight it is surprising, given the wide distribution of *B. garinii* and the apparent overlap of terrestrial and seabird cycles in Europe (Comstedt et al., 2006)
that in North America, *B. garinii* has not spread into inland areas and remains limited to coastal regions of Newfoundland (Smith et al., 2006). However, the lack of tick vectors that could maintain terrestrial transmission cycles in this region is likely a major reason why the seabird cycles have not spilled over into the rest of North America (Ogden et al., 2009a).

Differences in *Borrelia* transmission cycles also exist at a much finer scale driven by ecological factors, habitat types and microclimate which may locally determine tick and host abundance (Eisen et al., 2006; Fingerle et al., 2004; Hubalek and Halouzka, 1997; Killilea et al., 2008; Piesman, 2002; Rauter and Hartung, 2005).

Of the named species, seven occur in North America including *B. andersoni*, *B. bissettii*, *B. californensis*, *B. carolinensis*, *B. americana*, *B. kurtenbachii* and *B. burgdorferi* (Figure 4). *B. bissettii* has been found in Colorado, Illinois, California, North Carolina and South Carolina where *I. spinipalpis*, *I. pacificus*, or *I. affinis* act as vectors (Bissett and Hill, 1987; Lin et al., 2003; Maggi et al., 2010; Maupin et al., 1994; Norris et al., 1999; Picken et al., 1995; Postic et al., 1998). Although it had been reported that *B. bissettii* can be transmitted by *I. scapularis* under experimental conditions (Oliver, 1996), *B. bissettii* has not been found in questing *I. scapularis*.

Similarly, *B. kurtenbachii* has been isolated from host-derived larvae and DNA has been isolated from one questing adult *I. scapularis* (Anderson et al., 1988; Ogden et al., 2011; Picken and Picken, 2000) but the species has rarely been found in *I. scapularis* dominated habitats in recent years (Gatewood et al., 2009; Hamer et al., 2007; Hoen et al., 2009; Ogden et al., 2011; Oliver et al., 2006). If generalist vectors are able to transmit under experimental conditions LB group species that are usually transmitted by endophilic vectors, the question arises, why does it not happen more
frequently in natural transmission cycles, why are these species not more widely
distributed and what limits their distribution?

The species with the widest distribution in North America is *B. burgdorferi*,
ranging from NE, to Upper Midwest (MW) and Western States. It also occurs in some
Southern States and Southern Canada, within the distribution ranges of *I. scapularis*,
*I. pacificus*, and *I. affinis*. Interestingly, in North Carolina, it was found predominantly
in *I. affinis* but not *I. scapularis* and occured sympatrically with *B. bissettii* (Maggi et
al. 2010). In the NE *B. burgdorferi* appears to be the only LB species transmitted by *I.
scapularis*. Climatic conditions impacting tick phenology may favour selection of
certain strains (Diuk-Wasser et al., 2006; Gatewood et al., 2009; Ogden et al., 2007).

Both, *B. burgdorferi* and *B. bissettii* have been recorded in Europe and North America
(Postic et al. 1998, Gern and Humair 2002). In Europe, *B. bissettii* has been mainly
described from human patients (Picken et al., 1996a; Picken et al., 1996b; Rudenko et
al., 2008) but has only rarely been found in questing *I. ricinus* ticks (Hulinska et al.,
2007). Curiously, human infection with *B. bissetti* in the USA has not been reported.

Continued study of field-collected samples will likely continue to increase the number
of known *Borrelia* species (Scott et al., 2010).

6. Population structure and dispersal patterns of LB species

Host specialization is an important factor in vector-borne disease, and
different vector-borne pathogens show varying levels and patterns of host
specialization. An accurate understanding of the epidemiology of many zoonoses can
only be achieved by considering the varied ecological adaptations of the pathogens,
particularly differences in host specificity (Dubska et al., 2009; Hanincova et al.,
2003a; Hanincova et al., 2003b; Hu et al., 1997; Hu et al., 2001; Huegli et al., 2002;
Kurtenbach et al., 2001; Taragel'ova et al., 2008). The variation in host specialization makes the LB group of spirochetes an ideal model to directly contrast the effects of host specialization on the distribution of pathogens. As ticks cannot move over large distances independently (Falco and Fish, 1991), it has been suggested that the spread of LB spirochetes is linked to the movement of their hosts (Kurtenbach et al., 2002).

In addition to being of public health importance, the delineation and monitoring of the geographic ranges of the different LB species also provides opportunities to examine in more general terms the role of host ecology in the epidemiology of vector-borne zoonoses.

6.1 Europe and Asia

MLSA on housekeeping genes has revealed differences in the level of geographic structuring of populations of LB species that are consistent with distribution patterns of their different vertebrate hosts (Vitorino et al., 2008; Vollmer et al., 2011).

Vitorino and co-authors (2008) investigated B. lusitaniae, a species that has been associated with lizards, from two geographic regions in Portugal (Mafra and Grandola), which are approximately 160 km apart and located north and south of Lisbon. A pronounced fine-scale phylogeographic population structure was observed where most strains from Mafra clustered separately from Grandola strains (Vitorino et al., 2008). The authors suggested that this distribution reflects the highly parapatric population structure of lizards on the Iberian peninsula (Paulo et al., 2008).

Vollmer and co-authors (Vollmer et al., 2011) tested the prediction that host movement determines spirochaete biogeography by characterising B. garinii, B. valaisiana, and B. afzelii from various sites in Europe (Great Britain, France, Germany, Latvia). MLSA of the rodent-associated species, B. afzelii, showed a
population structure that signified restricted movement of strains between geographic regions. This differentiation was pronounced: only two B. afzelii STs have been found in more than one geographic location (Figure 5, Panel C). These data suggested that the English Channel may act as a barrier to the movement of B. afzelii strains between Great Britain and continental Europe (Vollmer et al., 2011). Chinese and European B. afzelii populations also showed high levels of differentiation suggesting very limited movement over these large distances. However, one Chinese B. afzelii strain clustered within the European group suggesting that there may be rare cases of movement between East and West, although the mechanisms behind such events are unclear (Vollmer, personal communication).

The data obtained by Vollmer and co-authors (Vollmer et al., 2011) are suggestive of interesting parallels between B. afzelii and the evolutionary history of their vertebrate host. In phylogenies and eBURST analyses, B. afzelii strains from Scotland appeared to be more closely related to STs found in Latvia than to STs found in England, suggesting that there is limited, or potentially no, movement of B. afzelii between north and south in the UK. This is interesting in the light of studies that investigated phylogenetic relationships of small mammals (including the field vole Microtus agrestis, bank vole Myodes glareolus, and pygmy shrew Sorex minutus) in Great Britain which showed a clear north/south divide between phylogroups (Searle et al., 2009). The marked differentiation between English and Scottish B. afzelii samples may therefore be a result of limited north-south rodent dispersal, although this hypothesis needs further investigation (Vollmer et al., 2011).

In addition, other studies of potential host species of B. afzelii including shrew and vole species have observed phylogeographic structuring of populations across Europe. These studies have attributed the phylogeographic patterns to population expansions
from ancestral refugia after the last glacial maximum (LGM), which possibly included
an Iberian and an East Baltic refuge (Heckel et al., 2005; Hewitt, 1999, 2001; Taberlet
and Bouvet, 1994; Taberlet et al., 1998). Northward spread of the populations from
the two refugia led to a possible overlap in the region of Germany or the Czech
Republic (Figure 6) (Heckel et al., 2005; Hewitt, 1999). Data of European B. afzelii
strains bear some resemblance of populations maintained potentially by the two
mammalian refuge populations as the B. afzelii phylogeny could be divided into a
Western European cluster and an Eastern European cluster (Vollmer, personal
communication).

However, fine scale structuring can also be observed in vole species either due
to natural and man made barriers (e.g. large rivers, highways) or due to a social
structure within population. These processes may limit the rates of movement
between host populations (Gerlach and Musolf, 2000; Schweizer et al., 2007) and
therefore limit dispersal of B. afzelii. Observations of B. afzelii strains at one site in
Latvia and the English sites are consistent with fine-scale structuring of the bacterial
populations due to restricted host movements (Vollmer et al., 2011). However, B.
afzelii has many rodent host species and further studies of small mammal host species
of B. afzelii may be required to better understand the ability of this species to disperse.

In contrast, both of the bird-related species investigated, B. valaisiana
and B. garinii, showed evidence of spatial mixing of STs between geographic regions
(Figure 5A, B) (Vollmer et al., 2011). Interestingly, while B. garinii data suggested
free movement of strains, B. valaisiana showed low to moderate differentiation,
suggesting there is not complete homogenization of B. valaisiana strains within
Europe. This was surprising because both species have been reported to be maintained
by similar species of avian hosts (Dubska et al., 2009; Taragel'ova et al., 2008) but
may suggest subtle ecological differences between these species. Certainly *B. garinii*

differs from *B. valaisiana* in being maintained in cycles between seabirds and their
associated tick, *I. uriae* (Bunikis et al., 1996; Larsson et al., 2007; Olsen et al., 1995;
Olsen et al., 1993) as well as in terrestrial cycles. Several studies (Comstedt et al.,
2006; Gomez-Diaz et al., 2011) reported an overlap of marine and terrestrial *B.
garinii* populations but the full impact on the observed population structure remains to
be investigated. Notably, *B. garinii* STs from China showed divergence from
European *B. garinii* STs indicated by long branches joining them to their closest
European relatives in phylogenetic inferences (Vollmer, personal communication) and
suggesting limited gene flow between the two regions. These data also suggested that
the role that migratory birds play in east-west or west-east movement of *B. garinii*
may be limited as would be expected as most migratory bird movement is on the
north-south axis. Analyses of more Russian and Asian *B. garinii* samples would be
required to confirm this hypothesis and assess the level of movement of *B. garinii*
between Asia and Europe.

Given that the movement of some LB species is limited by the propensity for
their vertebrate hosts’ ranges to shift, landscape genetic analysis would be an
appropriate approach to determine barriers to movement (Manel et al., 2003). Such
future investigations will be facilitated by identifying the full host spectrum of the
different LB species.

6.2 North America

In North America a complex picture of LB group species has emerged. While habitats
in California and the Southeastern States harbour a great variety of LB species, the
prevalence of human infections is low (Bacon et al., 2008). This may be related to
host preferences and human biting behaviour of main vectors in these regions or other ecological factors, although it may also be due to some of these species being non-pathogenic in humans (Eisen et al., 2004; Eisen et al., 2009; Girard et al., 2011; Lane and Quistad, 1998; Norris et al., 1996; Oliver, 1996; Oliver et al., 2003; Piesman, 1993; Swei et al., 2011; Talleklint-Eisen and Eisen, 1999; Wright et al., 1998). In Southeastern States, infection prevalence in I. scapularis ticks is low, which may be due to ‘dilution’ of transmission cycles by reservoir-incompetent lizards acting as tick hosts, and or by climate-driven tick seasonality that is less favourable for transmission (Durden et al., 2002; Ogden et al., 2008a; Ogden and Tsao, 2009b; Spielman et al., 1984); (Kollars et al., 1999; Spielman et al., 1984; Swanson and Norris, 2007).

In the NE USA, B. burgdorferi is the predominant species, human infection incidence is high (>80 % of all recorded infections in the USA occur here), and, here the first population level studies on B. burgdorferi were conducted. Pioneering studies using ospA and ospC as genetic markers (Qiu et al., 1997; Qiu et al., 2002; Wang et al., 1999) found a high local variation of strains but a uniform distribution across the NE USA. The authors suggested that ancient polymorphisms combined with balancing selection (in form of negative frequency-dependent immune selection) maintains the high diversity of populations (Dykhuizen et al., 1993; Qiu et al., 2002; Wang et al., 1999). Parallel studies on I. scapularis populations suggested that migration could also be at play as ‘American clade’ I. scapularis (Norris et al., 1996) were found in coastal bird sanctuaries in North Carolina (Qiu et al., 2002). Further studies on ospC including European and American strains of B. burgdorferi led the authors to suggest that recent and rapid spread of B. burgdorferi across two continents has occurred (Qiu et al. 2008). Transportation of ticks by infected migratory birds has more recently been suggested for the introduction of B. burgdorferi strains and I.
*scapularis* ticks into southern Canada (Ogden et al., 2010; Ogden et al., 2008b; Ogden et al., 2011; Ogden et al., 2006). Although both balancing selection and migration may have a homogenizing effect, there are several lines of evidence supporting the argument for balancing selection and/or functional constraints (related to its role in tissue adherence or protein binding during invasion/infection processes) acting on *ospC*: 1) identical *ospC* major types are found in all *B. burgdorferi* populations but these are regionally matched with different MLST STs (Margos et al., 2008; Qiu et al., 2008; Travinsky et al., 2010). This suggests that slowly evolving housekeeping genes have accumulated mutations while the *ospC* gene has not. 2) The description of *ospC* types that are found exclusively in Europe (e.g. P, Q, S, V) or California (e.g. H3, E3) points to population separation as frequent exchange between the populations would homogenize alleles (Girard et al., 2009; Qiu et al., 2008; Wang et al., 1999). The finding of these ‘private’ *ospC* types has been interpreted as adaptation to new habitats (Girard et al., 2009; Qiu et al., 2008). It could - alternatively - reflect loss of *ospC* major types in some regions due to severe population bottlenecks as described for the USA (Spielman, 1994).

MLST data based on housekeeping genes paint a different picture for *B. burgdorferi* populations. These data support the view that the *B. burgdorferi* populations from Europe and North America and in North America are genetically related but are currently separated with no or limited gene flow between them (Hoen et al., 2009; Margos et al., 2008; Ogden et al., 2011). In 2004 a CDC project was launched to investigate the presence and infection prevalences of *I. scapularis* nymphs on a country-wide scale and questing nymphs were collected systematically from May to September (Diuk-Wasser et al., 2006). Hoen and co-authors (Hoen et al., 2009) investigated the population structure of *B. burgdorferi* by MLST using 78
samples from 2004 and 2005: 41 samples were from NE sites and 37 from MW sites.

Thirty seven distinct STs were determined but no single ST was found in both regions suggesting restricted present day gene flow between the two regions. It further suggested that the coincident emergence of Lyme borreliosis in the two regions originated from multiple expansions of vector tick and B. burgdorferi populations (Hoen et al., 2009). Although the observed level of sequence divergence in some samples from NE and MW was only few nucleotides, considering the slow evolution of housekeeping genes, these mutational changes may have accumulated over time periods that exceeded the latest Lyme disease emergence in North America in the past 40 years. These results were supported by additional studies suggesting limited gene flow between the B. burgdorferi populations described from the NE and Upper MW ((Ogden et al., 2011), Bent, personal communication).

In a follow–up study, two transects in the NE were intensely sampled for questing I. scapularis nymphs in 2007, one transect starting from Long Island following the Hudson River valley north to Lake Champlain in Vermont, the other one starting in Old Lyme going north into Massachusetts. Although most STs were found in all sites, regression analysis revealed differences in frequencies of ST which correlated with latitude (Bent, personal communication). The mechanism behind the observed structure is currently unknown but possible reasons could be differences in host composition, pattern of movement, or genetic drift. Genetic drift can prevail and weaken natural selection in populations with small effective population size (N_e) ((Page and Holmes, 1998). Due to their parasitic life style, LB spirochetes are thought to have a small N_e (Qiu et al., 2002). Interestingly, although more than 300 samples were analysed, the number of new STs was small: four new STs were found in addition to the 37 STs described by Hoen et al. (Hoen et al., 2009) from the NE.
These data are consistent with limited genetic diversity in this region of high LB incidence probably reflecting the severe bottleneck that has been suggested previously (Spielman, 1994). Similar studies on I. scapularis samples collected by passive surveillance in Canada (from Nova Scotia to Manitoba) also supported the notion of geographic separation with limited gene flow between NE and MW. STs determined east of 80° longitude resembled those of NE USA, while STs west of 80° longitude resembled those of MW B. burgdorferi populations. Geographic analysis of STs and ospC alleles were consistent with south-to-north dispersion of infected ticks from the USA, likely on migratory birds (Ogden et al., 2011). Surprisingly, 19 novel STs were determined which were single (SLV), double (DLV) or triple locus variants (TLV) of STs from the USA supporting the notion that the spatial scale of sampling is important to capture the population variation of, and to understand demographic processes in, B. burgdorferi (Figure 7). Preliminary MLST data from approximately 25 strains from the Upper MW and 25 strains from California show that additional samples from these regions led to denser eBurst ‘forests’ and better resolution of clonal complexes. Indeed, in the Californian dataset the first SLV of ST1 (B31) was determined (Margos, unpublished) supporting the view that B. burgdorferi populations across North America – not only in the MW and NE but also from California – are genetically related and once belonged to an admixed population (Hoen et al., 2009).

The complexity observed for B. burgdorferi populations in North America is likely due to a dynamic short- and long-term evolution. The long-term evolutionary history was probably shaped by glacial-interglacial cycles (Humphrey et al., 2010; Qiu et al., 2002) which is consistent with data by Hoen et al. (2009) who found signatures of ancient population expansions of B. burgdorferi likely to date back
several thousand, if not millions of years ago. Demographic events in the past 200 years (following the arrival of European settlers) have shaped populations of hosts and vectors by deforestation, dwindling deer and tick populations and causing severe bottlenecks in *Borrelia* populations (McCabe and McCabe, 1997; Spielman, 1994). Since then, expansion of deer and tick populations have resulted in the latest dispersal of LB spirochetes leading to an epidemic of human LB in the NE and MW USA during the past four decades (Bacon et al., 2008). It is conceivable that different regions were affected in different ways by these processes but in order to understand the contemporary pattern sequence data with high resolution power, such as genome wide SNPs, will be required (Figure 9).

*B. burgdorferi* is considered a generalist species that can be maintained and transmitted to ticks by a great variety of hosts including birds and rodents (Brisson and Dykhuizen, 2006; Hanincova et al., 2006; Richter et al., 2000), therefore, understanding its dispersal is more complicated than that of host specialized species. Fitness variation in hosts has been described for several strains (Derdakova et al., 2004; Hanincova et al., 2008b) and host adaptations may be developing (Brinkerhoff et al., 2010; Brisson and Dykhuizen, 2004; Ogden et al., 2011); all of which is likely to impact transmission efficiency and dispersal of *B. burgdorferi* strains. Some models of dispersal of *B. burgdorferi* have emerged that are consistent with slow south-to-north range expansions of *B. burgdorferi* that lag behind expansion of the tick vector (Ogden et al., 2010). Whether or not there is low level east-west or west-east migration, is far less understood (Hamer et al., 2010; Ogden et al., 2011). Clearly, additional information on the ecology of *B. burgdorferi* strains is required in order to obtain a comprehensive picture of how *B. burgdorferi* strains spread.
7. Models of global evolution

For phylogenetic analyses of the whole group of LB spirochetes, housekeeping genes provide the benefit of defining outgroup species as they are also present in the relapsing fever spirochetes (B. hermsii, B. duttoni, B. turicatae) allowing rooting of phylogenies. Their analysis also allows inferences of the temporal evolution of LB species. However, ascertaining this order using a single gene such as flaB, or even MLSA, proved difficult due to low confidence values of internal branches in phylogenies in which all STs of the European species were included (Fukunaga et al., 1996c; Kurtenbach et al., 2010). Several factors may be responsible for this: 1) internal branches representing species divisions are extremely short suggesting that the speciation events, in evolutionary terms, occurred in quick succession. Thus there are limited mutations existing in the sequences today that represent these intermediate species. 2) The limited number of genes may not contain a sufficient number of nucleotide polymorphisms to clearly define the topology. 3) These short branches may be suggestive of incomplete lineage sorting (Avise and Robinson, 2008; Maddison and Knowles, 2006). This occurs when polymorphisms are maintained in a gene through two or more speciation events, thus giving the impression of a different topology.

Several unrooted or midpoint rooted phylogenies have been published for LB group species (Margos et al., 2010; Richter et al., 2006; Rudenko et al., 2009b) (Figure 8). These trees produced different topologies compared to each other and to the concatenated housekeeping gene trees. Differences in tree topology are most notable comparing phylogenetic inferences for MLSA genes and ospC suggesting different evolutionary pathways of plasmid encoded and chromosomal genes (Figures 2A, B).
There are, however, some species that form clusters in all trees generated using single or multiple chromosomal loci. Notably, the ‘American’ species and the ‘Eurasian’ species form sister clades joined by a well supported branch suggesting that these two clades separated early during LB evolution. Within the ‘American’ clade, \textit{B. burgdorferi} and \textit{B. bissettii} (both occurring in North America and Europe) fall into different subclades raising questions about migration times and routes between continents. Several species (if included in phylogenies) tend to always cluster closely together such as \textit{B. afzelii} and \textit{B. spielmanii} or \textit{B. garinii} and \textit{B. bavariensis} being consistent with more recent speciation events (Margos et al., 2010; Postic et al., 2007; Rudenko et al., 2009b). It is also apparent that host associations did not develop only once. For example, not all bird-adapted LB species cluster monophyletically in the species tree (Figure 2A) suggesting that several host switches occurred during the evolutionary history of LB species.

The doubling time of LB spirochetes in feeding nymphs has been estimated to be four hours but was much slower \textit{in vitro}, approximating 8-12 h under constant temperature conditions (33°C) (De Silva and Fikrig, 1995; Heroldova et al., 1998; Pollack et al., 1993) and may be even longer at lower (winter) temperatures in vector populations under natural conditions. It is, therefore, extremely difficult to estimate mutation rates or time of speciation events for LB species by comparison with other bacterial species. Consequently, to establish a realistic time frame of LB species evolution, measures of mutation rates for LB species are essential which will require the use of larger sets of sequence data than MLSA.

\textbf{8. Future Avenues}
In this paper we have summarized recent research on population genetics, molecular
taxonomy and evolution of LB spirochetes which has moved from single locus
approaches to multilocus approaches. From the information gathered here, it is
evident that major advances have been made in understanding the evolutionary
ecology of LB spirochetes but there are also limitations which need to be addressed.
These include questions like: 1) What drives associations/adaptation between LB
group spirochetes and their hosts and vectors? 2) What is the full host spectrum of the
different LB species? 3) Which factors apart from host associations impact dispersal
of LB group species? 4) What is the speed and geometry of spread?

Some methods have been developed to address such questions. For example
blood meal analyses analysis in questing ticks may help to resolve host associations
(Humair et al. 1997) and real time genotyping assays are already being used for LB
spirochetes for loci such as IGS, fla or hbb (Herrmann and Gern, 2010; Portnoi et al.,
2006; Strube et al., 2010), but these methods may need refinement. Next generation
sequencing and SNP analyses will likely proof very valuable to develop better tools
for precise strain identification, to identify mixed infections in ticks or patients, to
refine blood meal analysis or to address questions regarding the deep evolutionary
relationships of LB group spirochetes. Developments such as single nucleotide primer
extension assays (Murphy et al., 2003) or high melting resolution techniques (Wittwer
et al., 2003) may be suitable for such approaches.

MLST of housekeeping genes in LB spirochetes has shown that recombination
can occur on chromosomally located loci. In general, the use of MLST/MLSA has
shown that there is great variation in bacterial inheritance. While some taxa such as
Staphylococcus aureus, Yersinia sp or Salmonella typhi show little horizontal gene
transfer, others show an enormous amount of recombination or horizontal gene
transfer, well known examples are *Neisseria* sp and *Helicobacter pylori* (Achtman, 2004; Feil et al., 2003; Feil and Spratt, 2001; Holt et al., 2008; Ochman et al., 2000).

However, the availability of whole genome sequences for a large numbers of closely related bacteria has led to the concept that most bacterial genomes consists of a ‘core’ genome which can inform about evolutionary relationships and an ‘accessory’ genome which is much more flexible, permits invasion of new niches or confers selective advantages (such as antibiotic resistance) by horizontally acquired genome elements (often plasmids) (Guttman and Stavrinides, 2010; Ochman et al., 2000).

For some microbial pathogens whole genome sequencing and detection of SNPs has led to a quantum leap forward in understanding evolutionary and demographic processes. Fortunately, the haploid nature of bacteria makes it convenient to detect these processes using sequence data. Genome-wide SNPs permit us to distinguish evolutionary processes, such as mutation, recombination, selection, and drift, from demographic processes affecting the whole genome such as migration, population expansion/contractions (Guttman and Stavrinides, 2010). Analyses of genome-wide SNPs permit much more accurate estimates of mutation rates which is a vital parameter in population level and evolutionary studies, as well as the analyses of recent as well as ancient events (Figure 9). Computer programs have been developed for bacterial population based inferences and many are tailored for use with MLST/MLSA or SNP data (Corander and Marttinen, 2006; Corander et al., 2004; Didelot et al., 2009; Didelot and Falush, 2007; Feil et al., 2004; Francisco et al., 2009; Guillot, 2008; Guillot et al., 2008; Kuhner, 2006; Schierup and Wiuf, 2010) (see review by (Excoffier and Heckel, 2006)).

Several *Borrelia* genomes have been sequenced and for *B. burgdorferi* more than 10 draft genomes are available (Schutzer et al., 2011). While this is a good start
and can provide the scaffolding for next generation sequencing of further samples, understanding the most recent population expansion in northeastern America requires the analyses of carefully selected samples from that region. As MLST and eBurst data convincingly demonstrate, getting insights into the deep evolutionary history of *B. burgdorferi* requires sampling at a different scale. In our opinion – the time is ripe to take *Borrelia* research to the next step, and that is the emerging field of bacterial population genomics (Guttman and Stavrinides, 2010) as this together with MLST will provide a framework for epidemiological, clinical and ecological studies.

**Acknowledgements**

We would like to thank S. J. Bent, J. Tsao and R. S. Lane for sharing unpublished data, numerous colleagues and the German Collection of Microorganisms and cell cultures (DSMZ) for providing *Borrelia* DNA and a large number of tick collectors for their efforts. The authors are grateful for financially support received by The Wellcome Trust (grant no. 074322/Z/04/Z), Public Health Agency of Canada, NIH-NIAID (grant nos. AR041511; 5R21AI065848-03); USDA-ARS Cooperative Agreement; G. Harold and Leila Y. Mathers Charitable Foundation; US Centers for Disease Control and Prevention.

**References**


ticks by using electrophoretic chips and real-time polymerase chain reaction. Folia Microbiol (Praha) 52, 315-324.


Lane, R.S., Quistad, G.B., 1998. Borreliacidal factor in the blood of the western fence lizard (Sceloporus occidentalis). J Parasitol 84, 29-34.


Ogden, N.H., Lindsay, L.R., Morshed, M., Sockey, P.N., Artsob, H., 2009a. The emergence of Lyme disease in Canada. CMAJ Canadian Medical Association Journal 180, 1221-1224.


constitute a key reservoir of *Borrelia garinii*, the causative agent of borreliosis in Central Europe. Appl Environ Microbiol 74, 1289-1293.


Table 1 List of putative and named species within the LB group spirochetes, their host and vector range and distribution.

<table>
<thead>
<tr>
<th>Species (c/p)* (type strain)</th>
<th>Distribution</th>
<th>Host range</th>
<th>main vector</th>
<th>References for spec. description</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. afzelii</em> (c) (VS461)</td>
<td>Europe, Asia</td>
<td>Apodemus spp, Myodes glareolus, Sorex spp, Sciurus spp, Erinaceus spp, Rattus spp</td>
<td>Ixodes ricinus, I. persulcatus, I. hexagonus</td>
<td>Canica et al. 1993</td>
</tr>
<tr>
<td><em>B. andersonii</em> (c) (21038)</td>
<td>North America</td>
<td>Sylvilagus spp, (Passeriformes spp)</td>
<td>I. dentatus</td>
<td>Marconi et al. 1995</td>
</tr>
<tr>
<td><em>B. bavariensis</em> (p) (PBi)</td>
<td>Europe, Asia (?)</td>
<td>Apodemus spp, Myodes sp, Microtus spp.</td>
<td>I. ricinus, I. persulcatus(?)</td>
<td>Margos et al. 2009</td>
</tr>
<tr>
<td><em>B. bissettii</em> (c) (DN127-cl9-2)</td>
<td>North America, Europe</td>
<td>Neotoma spp, Peromyscus spp, Sigmodon spp</td>
<td>I. pacificus, I. spinipalpis, I. affinis, EU: unknown</td>
<td>Postic et al. 1998</td>
</tr>
<tr>
<td><em>B. californiensis</em> (c) (CA446)</td>
<td>Western US</td>
<td>Dipodomys californicus</td>
<td>unknown</td>
<td>Postic et al. 2007</td>
</tr>
<tr>
<td><em>B. carolinensis</em> (c) (SCW-22)</td>
<td>Southeast US</td>
<td>P. gossypinus, N. floridana</td>
<td>unknown (I. minor?)</td>
<td>Rudenko et al. 2009a</td>
</tr>
<tr>
<td><em>B. garinii</em> (c) (20047)</td>
<td>Europe, Asia, Artic-Antartic circles</td>
<td>Turdus merula, T. philomelos, Parus major, seabirds (Puffin, Guillemot, Kittiwake, Razorbill)</td>
<td>I. ricinus, I. persulcatus, I. uriæ</td>
<td>Baranton et al. 1992</td>
</tr>
<tr>
<td><em>B. japonica</em> (c) (HO14)</td>
<td>Japan</td>
<td>Sorex unguiculatus, Apodemus spp, Eothenomys smithi</td>
<td>I. ovatus</td>
<td>Kawabata et al. 1993, Postic et al. 1993</td>
</tr>
<tr>
<td><em>B. kurtenbachii</em> (p) (25015)</td>
<td>Northamerica, (Europe?)</td>
<td>Microtus pennsylvanicus, Zapus hudsonius, Peromyscus?</td>
<td>unknown (I. scapularis?)</td>
<td>Margos et al. 2010</td>
</tr>
<tr>
<td><em>B. lusitaniae</em> (c) (PoTIB2)</td>
<td>Mediterranean basin</td>
<td>Lacertidae</td>
<td>I. ricinus</td>
<td>LeFleche et al. 1997</td>
</tr>
<tr>
<td><em>B. sinica</em> (c) (CMN3)</td>
<td>China</td>
<td>Niviventer confucianus</td>
<td>I. ovatus</td>
<td>Masuzawa et al. 2001</td>
</tr>
<tr>
<td><em>B. spielmanii</em> (c)</td>
<td>Europe</td>
<td>Glis glis,</td>
<td>I. ricinus</td>
<td></td>
</tr>
</tbody>
</table>

* c: competent host, p: partial host

---

53
<table>
<thead>
<tr>
<th>(PC-Eq17N5)</th>
<th><em>Eliomus quercinus</em></th>
<th>Richter et al. 2006</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>B. tanukii</em> (c) (Hk501)</td>
<td>Japan</td>
<td><em>Apodemus</em> sp, <em>Clethrionomys rufocanus</em>, <em>Eothenomys smithii</em></td>
</tr>
<tr>
<td><em>B. turdi</em> (c) (Ya501)</td>
<td>Japan</td>
<td><em>Turdus</em> spp</td>
</tr>
<tr>
<td><em>B. valaisiana</em> (c) (VS116)</td>
<td>Europe, Japan</td>
<td><em>Turdus merula</em>, <em>T. philomelos</em>, <em>Parus major</em></td>
</tr>
<tr>
<td><em>B. yangtze</em> (p) (nd)</td>
<td>China</td>
<td><em>Niviventer fulvescens</em>, <em>Apodemus</em> sp</td>
</tr>
<tr>
<td>Genomospecies2</td>
<td>United States</td>
<td>unknown</td>
</tr>
</tbody>
</table>

\(^{a}c – confirmed; p – proposed; nd = not determined\)
**Table 2** Typing schemes for LB spirochetes using multiple loci

<table>
<thead>
<tr>
<th>Type of Loci</th>
<th>Loci</th>
<th>purpose</th>
<th>data</th>
<th>reference</th>
</tr>
</thead>
<tbody>
<tr>
<td>17 plasmid-encoded loci, chromosomal: housekeeping gene</td>
<td>lp54, cp26, cp9, lp17, lp25, lp28-2, lp28-4, lp38, BB0082</td>
<td>population studies</td>
<td>GenBank, ~60 strains</td>
<td>Qiu et al. 2004</td>
</tr>
</tbody>
</table>
Figure 1. Factors impacting the evolutionary ecology of LB spirochetes. Biotic factors are shown next to the host-vector-spirochete triangle. Abiotic factors (such as climate or landscape) act indirectly on LB spirochetes by impacting on host and vector populations. The contemporary picture is further compounded by the evolutionary and demographic history of hosts, vectors, and pathogens.
Figure 2. Bayesian phylogenetic inferences generated using MLST housekeeping genes (A) and ospC (B) sequences. Previously assigned species are color coded as follows: *B. burgdorferi* s.s. – ■, *B. afzelii* – □, *B. garinii* – ●, *B. bavariensis* - △.
B. valaisiana – ○, and B. lusitaniae – ▼. The MLST tree was rooted with sequences of the relapsing fever spirochetes B. duttonii, B. hermsii, and B. turicatae. The branch length of the outgroup is not according to scale as indicated by slashes. While in the MLST tree LB species cluster monophyletically, this is not the case using ospC sequences (original figure A from Population Biology of Lyme Borreliosis Spirochetes; Kurtenbach et al [2010]. DOI: 10.1002/9780470600122.ch12; Copyright (2010, John Wiley & Sons); reprinted with permission of John Wiley & Sons, Inc.; original figure B Margos et al. [2009], doi 10.1128/AEM.00116-09, reproduced and modified with permission from the American Society for Microbiology)
Figure 3. Multi Locus Sequence Typing. Targeted PCR is used to amplify several genes distributed throughout the genome exhibiting nearly neutral variation. Internal fragments, kept in-frame, of similar length for each gene are used. For each individual gene, fragments of identical length are aligned and compared to sequences in a ‘virtual strain collection centre’, a MLST database, and to each other permitting determination of an allelic profile for each strain. The allelic profile determines the sequence type (ST) and it can be used to infer relationships of descent within bacterial species based on models of clonal expansion and diversification. Concatenated sequences of all genes can be used for phylogenetic inferences. The accumulative nature of MLST database makes it an attractive instrument to understand intra- and inter-specific relationships of bacteria.
Figure 4. Map showing the global distribution of the LB species. The shaded areas show the distribution of tick vectors. Seven species of LB group spirochetes are found in North America, eight species in Europe, and eight species in Asia, two species overlap in the Old and New Worlds, three in Europe and Asia (see text for details).
Figure 5. goeBURST diagrams based on the multi-locus allelic profiles for *B. garinii* (A), *B. valaisiana* (B) and *B. afzelii* (C). Each coloured box represents an ST. The colour and size of the boxes corresponds to geographic region and the number of that ST found. STs unique to a particular country were coloured as follows: red England, blue France, yellow Germany, green Latvia, purple Scotland. Those STs that were found in more than one country are grey. STs connected by black or blue lines are single-locus variants (SLVs) and STs connected by grey or green lines are double-locus variants (DLVs) (original figure from Vollmer et al. [2011] *Environmental Microbiology*, doi:10.1111/j.1462-2920.2010.02319.x, reproduced with permission from John Wiley and Sons)
Figure 7. A population snapshot of 244 samples of *Borrelia burgdorferi* found in Canada (166 samples) and the United States (78 samples) as determined by spatial analysis using spatialepidemiology.net. The figure reveals correspondence of sequence type and geographic distribution. Most ST were found either in the Northeast or the Midwest suggesting limited gene flow between populations.
Figure 8 Neighbour joining tree generated using concatenated sequences of MLSA housekeeping genes showing LB groups species. Black dots indicate species that occur in North America, circles indicate species that occur in Eurasia, grey dots indicate species that occur in the Old and New Worlds. The scale bar shows 1% divergence. Branch confidence values calculated using a bootstrap procedure with 100 repeatitions (original figure from Margos et al. [2010] Ticks and Tick-borne Diseases, doi: 10.1016/j.ttbdis.2010.09.002, modified and reproduced with permission from Elsevier)
Figure 9 Graphic representation of the ‘time’ captured by various genetic elements used for typing of bacterial microorganisms. The highly conserved 16S locus reveals deep evolutionary relationships but is unable to capture recent events. Fast evolving genetic elements, such as loci under diversifying selection, microsatellites or variable number of tandem repeats (VNTR) may reveal very recent events but – due to saturation – are not able to ‘see’ ancient events. Intergenic spacer (IGS) regions are supposed to be selectively neutral and should therefore accumulate mutations indiscriminately and linear to time. IGS may be short and saturate quickly or may contain regulatory elements which might not permit all mutations to be fixed. Due to the slow evolution of housekeeping genes multilocus sequence typing captures the intermediate relationship of bacteria. Genome-wide SNPs provide the broadest ‘view’ on an organism past as these are able to capture recent as well as ancient events.