Review

Population genetics, taxonomy, phylogeny and evolution of *Borrelia burgdorferi* sensu lato

Gabriele Margos,⇑ Stephanie A. Vollmer, Nicholas H. Ogden, Durland Fish

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

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

*Yale School of Public Health, New Haven, CT 06520, United States*

**Article info**

**Article history:**
Received 12 May 2011
Received in revised form 29 July 2011
Accepted 31 July 2011
Available online 5 August 2011

**Keywords:**
*Borrelia burgdorferi* Evolution Molecular ecology Ixodes Multilocus sequence typing MLST

**Abstract**

In order 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.

© 2011 Elsevier B.V. All rights reserved.

**Contents**

1. Introduction ................................................................. 1545
2. Ecology of LB group of spirochetes .................................. 1547
3. Typing tools for the LB group of spirochetes ....................... 1548
   3.1. Interspecies studies ................................................. 1548
   3.2. Intraspecies genotyping ........................................... 1548
   3.3. Typing schemes using multiple loci .......................... 1550
4. Borrelia taxonomy ......................................................... 1552
5. Geographic distribution .................................................. 1552
6. Population structure and dispersal patterns of LB species ..... 1553
   6.1. Europe and Asia ................................................... 1553
   6.2. North America ..................................................... 1555
7. Models of global evolution .............................................. 1557
8. Future avenues ........................................................... 1557
Acknowledgements .......................................................... 1558
References ................................................................. 1558

Abbreviations: IGS, intergenic spacer; LB, Lyme borreliosis; MLST/MLSA, multilocus sequence typing/multilocus sequence analysis; MW, Midwest; NE, Northeast; Osp, outer surface protein; SLV/DLV/TLV, single locus variant/double locus variant/triple locus variant; SNP, single nucleotide polymorphism; ST, sequence type.

⇑ Corresponding author. Address: Department of Biology and Biochemistry, 3 South, University of Bath, Claverton Down, Bath BA2 7AY, UK. Tel.: +44 1225 385116; fax: +44 1225 386779.
E-mail address: gm250@bath.ac.uk (G. Margos).

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
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,b).

### Table 1

<table>
<thead>
<tr>
<th>Species (c/p)*</th>
<th>Distribution</th>
<th>Host range</th>
<th>Main vector</th>
<th>References for spec. description</th>
</tr>
</thead>
<tbody>
<tr>
<td>B. afzelii (c)</td>
<td>Europe, Asia</td>
<td>Apodemus spp., Myodes glareolus, Sorex spp., Sciurus spp., Erinaceus spp., Rattus spp.</td>
<td>l. persicinus, l. persulcatus, l. hexagonus</td>
<td>Canica et al. (1993)</td>
</tr>
<tr>
<td>B. americana (p)</td>
<td>North America</td>
<td>Thryothorus ludovicianus, Pipilo erythrophthalmus</td>
<td>l. pacificus, l. minor</td>
<td>Rudenko et al. (2009b)</td>
</tr>
<tr>
<td>B. andersonii (c)</td>
<td>North America</td>
<td>Sylvilagus spp., (Passeriformes spp.)</td>
<td>l. dentatus</td>
<td>Marconi et al. (1995)</td>
</tr>
<tr>
<td>B. bavariensis (p)</td>
<td>Europe, Asia</td>
<td>Apodemus spp., Myodes sp., Microtus spp.</td>
<td>l. ricinus, l. persulcatus (?), l. pacificus, l. spinipalpis, l. affinis, EU: unknown</td>
<td>Margos et al. (2009)</td>
</tr>
<tr>
<td>B. bissetti (c)</td>
<td>North America, Europe</td>
<td>Neotoma spp., Peromyscus spp., Sigmodon spp.</td>
<td>l. ricinus, l. hexagonous, l. scapularis, l. pacificus, l. affinis, l. minor, l. spinipalpis, l. muris</td>
<td>Postic et al. (2007)</td>
</tr>
<tr>
<td>B. californiensis (c)</td>
<td>Western US</td>
<td>Dipodomys californensis</td>
<td>Unknown</td>
<td>Postic et al. (2007)</td>
</tr>
<tr>
<td>B. carolinensis (c)</td>
<td>Southeast US</td>
<td>P. gossypinus, N. floridana</td>
<td>Unknown</td>
<td>Rudenko et al. (2009a)</td>
</tr>
<tr>
<td>B. garinii (c)</td>
<td>Europe, Asia, Artic-Antarctic circles</td>
<td>Tardus merula, T. philomelos, Parus major, seabirds (Puffin, Guillemot, Kittiwake, Razorbill)</td>
<td>l. ovatus</td>
<td>Kawabata et al. (1993), Postic et al. (1993)</td>
</tr>
<tr>
<td>B. japonica (c)</td>
<td>Japan</td>
<td>Sorex ungulatus, Apodemus spp., Eothenomys smithi</td>
<td>l. ovatus</td>
<td>Margos et al. (2010)</td>
</tr>
<tr>
<td>B. kentucky (p)</td>
<td>Northamerica, (Europe?)</td>
<td>Microtus pennsylvanicus, Zapus hudsonius</td>
<td>Unknown</td>
<td>Le Fleche et al. (1997)</td>
</tr>
<tr>
<td>B. lusitaniae (c)</td>
<td>Mediterranean basin</td>
<td>Lacteriae</td>
<td>l. ricinus</td>
<td>Masuzawa et al. (2001)</td>
</tr>
<tr>
<td>B. sinnia (c)</td>
<td>China</td>
<td>Niviventer confluicanus</td>
<td>l. ovatus</td>
<td>Richter et al. (2006)</td>
</tr>
<tr>
<td>B. spielmani (c)</td>
<td>Europe</td>
<td>Glis glis, Eliomus quercinus</td>
<td>l. ovatus</td>
<td>Fukunaga et al. (1996a-c)</td>
</tr>
<tr>
<td>B. tanuki (c)</td>
<td>Japan</td>
<td>Apodemus sp., Clethrionomys rufocanus, Eothenomys smithii</td>
<td>l. tanuki</td>
<td>Fukunaga et al. (1996a-c)</td>
</tr>
<tr>
<td>B. tundi (c)</td>
<td>Japan</td>
<td>Tardus spp.</td>
<td>l. tundus</td>
<td>Wang et al. (1997)</td>
</tr>
<tr>
<td>B. valaisiana (c)</td>
<td>Europe, Japan</td>
<td>Tardus merula, T. philomelos, Parus major</td>
<td>l. tundus, l. ricinus, l. columnae</td>
<td>Wang et al. (1997)</td>
</tr>
<tr>
<td>B. yangtze (c)</td>
<td>China</td>
<td>Niviventer fulvescens, Apodemus sp.</td>
<td>l. granulatus, l. nipponensis, l. spinipalpis, l. pacificus</td>
<td>Postic et al. (2007)</td>
</tr>
<tr>
<td>Genomospecies2</td>
<td>United States</td>
<td>Unknown</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

* c – confirmed; p – proposed; nd = not determined.
(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, *Borrelia afzelii* is most frequently linked with skin manifestations, *Borrelia garinii* and *Borrelia bavariensis* with neuroborreliosis, and *B. burgdorferi* with arthritic symptoms (Canica et al., 1993; Ornstein et al., 2001; Randolph, 2008; Rijkema et al., 1997; Stanek and Strle, 2009; Steere et al., 1986; van Dam, 2002). Some species, such as *Borrelia lusitaniae*, have only occasionally been associated with human disease while for others, such as *Borrelia 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 infection. LB species vary in their geographic distribution, host specificity and ability to cause disease in humans (Baranton et al., 2001; Seinost et al., 1999; Wilske 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 (Fig. 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 individual host, although not necessarily a different host species. Except in the case of nicoliodic (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 their natural transmission cycles. The bacteria are taken up during a bloodmeal from an infected and infectious host, are maintained transstadially during the molting 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 *Ixodes scapularis* and *Ixodes persulcatus* (Nefedova et al., 2004; Patrican, 1997). However, relapsing-fever like spirochetes (e.g. *Borrelia 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 *I. 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; Love and Lane, 1988; Xu et al., 2003). These are *Ixodes ricinus* in Europe, *I. persulcatus* in Eastern Europe and Asia, *I. scapularis* and *Ixodes pacificus* in North America. Other experimentally confirmed vector-competent *Ixodes* species are nicoliodic 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 nicoliodic 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), racoons 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 (Mathet 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., 1998b; 2002a,b; 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., 2003a,b; Hu et al., 1997, 2001; Humair et al., 1998, 1999; Kurtenbach et al., 1998a; Tarageł’ova et al., 2008). 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, such hosts 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). (3) Recent evidence supports the view that host associations substantially shape Borrelia populations by impacting their dispersal 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 hosts 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 (Fig. 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 (Qi et al., 2002)), this may weaken natural selection and introduce stochastic effects into allele frequencies of populations (Page and Holmes, 1998). The signature of 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; Wilks et al., 1991). Phenotypic typing tools were developed to reveal intraspecific diversity which included serotyping or multilocus enzyme electrophoresis (MLEE) (Boerlin et al., 1992; Wilks et al., 1991, 1993, 1995, 1996). This topic is reviewed excellently by 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 recombination 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; Wilks 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 thus 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 \((rsl–rta)\) large subunits (Schwartz et al., 1992). The IGS between the 5S and 23S \((rfl–rfl)\) 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, 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 (Rijksema 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., 1998a, 2001). 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 \((rsl–rta)\) IGS or outer surface protein (Ospa) encoding loci for B. burgdorferi in North
Fig. 2. Bayesian phylogenetic inferences generated using MLST housekeeping genes (A) and ospC (B) sequences. Previously assigned species are 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); original figure B Margos et al. (2009), reproduced and modified with permission from the American Society for Microbiology).
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) (Fig. 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 was found (Hellingren 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; Vittorino et al., 2008; Wang et al., 1999a unpublished). While 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 (Vittorino 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., 1999c; 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 (Fig. 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 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 www.mlst.net (Aanensen and Spratt, 2005). 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., 2005, 2006). 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 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., 2009, 2010; Postic et al., 2007; Richter et al., 2006; Rudenko et al., 2009a,b). 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

**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>References</th>
</tr>
</thead>
</table>
evolutionary inferences. For the MLST 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 1200 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). MLSA, the genome-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 H014, 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 Fig. 2A).

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. spielmani, B. californiaensis, B. carolinensis, B. americana, B. yangtze, B. bavariensis and B. turdenbachii) by genetic distance analyses (Chu et al., 2008; Margos et al., 2010, 2009; Postic et al., 2007; Richter et al., 2006; Rudenko et al., 2009a,b). Richter et al. (2006) and 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 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 strains 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 (Fig. 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 k语音 whose columns 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 I语音 vector adaptations of Borrelia species. Species with a localised distribution are B. turdi, B. denticola, 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; Yousui 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; Wodecka 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 (Dunneau 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., 2009).

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; Killiea et al., 2008; Piesman, 2002; Rauter and Hartung, 2005).

Of the named species, seven occur in North America including B. andersonii, B. bissetti, B. californiaensis, B. carolinensis, B. americana, B. turdenbachii and B. burgdorferi (Fig. 4). B. bissetti 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., 2011).
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* 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?

6. Population structure and dispersal patterns of LB species

Host specialization is an important factor in vector-borne diseases, 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,b; Hu et al., 1997, 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 geographic 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., 2002c). 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 et al. (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 spirochete 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 (Fig. 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).
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 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 *Mystes glareolus*, and pygmy shrew *Sorex minutus*) in Great Britain which show 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 (LCM), 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 (Fig. 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* phylogeeny 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 (Fig. 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; Taragefova 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, 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

---

**Fig. 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.1462-2920.2010.02315.x, reproduced with permission from John Wiley and Sons).
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, 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, 2009; 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, 2002; Wang et al., 1999c) 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., 1999c). 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). Transmission 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., 2008b, 2006, 2010, 2011). 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) points to population separation as frequent exchange between the populations would homogenize alleles (Girard et al., 2009; Qiu et al., 2008; Wang et al., 1999c). 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 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.

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* (Fig. 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 (Fig. 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

---

**Fig. 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 STs were found either in the Northeast or the Midwest suggesting limited gene flow between populations.
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 informs of the temporal evolution of LB species. However, ascertaining this order using a single gene such as *fla*β, 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) (Fig. 8). These trees produced different topologies compared to each other and to the concatenated housekeeping gene trees. Differences in tree topology are most topologies compared to each other and to the concatenated sequences of MLSA allowing rooting of 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.

There are, however, some clusters 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, *B. burgdorferi* and *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 *B. afzelii* and *B. spielmanii* or *B. garinii* and *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 (Fig. 2A) suggesting that several host switches occurred during the evolutionary history of LB species.

The doubling time of LB spirochetes in feeding nympha has been estimated to be four hours but was much slower 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 would be considerably 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.

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 the associations/adaptation between LB group spirochetes and their hosts and vectors? (2) What is the full host spectrum of the differ-
ent 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 in questing ticks may help to resolve host associations (Humair et al., 2007) 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 of evolutionary relationships of LB group spirochetes. Developments such 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. We 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


