Microbiology (2004), **150**, 1741–1755 DOI 10.1099/mic.0.26944-0

Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe

Jonas Bunikis,[1] Ulf Garpmo,[2]† Jean Tsao,[3,4]† Johan Berglund,[5] Durland Fish[3] and Alan G. Barbour[1]

[1]Departments of Microbiology and Molecular Genetics and Medicine, B240 Medical Sciences I, University of California Irvine, Irvine, CA 92697-4025, USA

[2]Kalmar County Hospital, Kalmar, Sweden

[3]Department of Epidemiology and Public Health, Yale University School of Medicine, New Haven, CT, USA

[4]Department of Fisheries and Wildlife, Michigan State University, MI, USA

[5]Department of Community Medicine, Lund University, Lund, Sweden

Correspondence
Jonas Bunikis
jbunikis@uci.edu

Alan G. Barbour
abarbour@uci.edu

Received 27 November 2003
Revised 23 February 2004
Accepted 5 March 2004

The genetic polymorphism of *Borrelia burgdorferi* and *Borrelia afzelii*, two species that cause Lyme borreliosis, was estimated by sequence typing of four loci: the *rrs–rrlA* intergenic spacer (IGS) and the outer-membrane-protein gene *p66* on the chromosome, and the outer-membrane-protein genes *ospA* and *ospC* on plasmids. The major sources of DNA for PCR amplification and sequencing were samples of the *B. burgdorferi* tick vector *Ixodes scapularis*, collected at a field site in an endemic region of the north-eastern United States, and the *B. afzelii* vector *Ixodes ricinus*, collected at a similar site in southern Sweden. The sequences were compared with those of reference strains and skin biopsy isolates, as well as database sequences. For *B. burgdorferi*, 10–13 alleles for each of the 4 loci, and a total of 9 distinct clonal lineages with linkage of all 4 loci, were found. For *B. afzelii*, 2 loci, *ospC* and IGS, were examined, and 11 IGS genotypes, 12 *ospC* alleles, and a total of 9 linkage groups were identified. The genetic variants of *B. burgdorferi* and *B. afzelii* among samples from the field sites accounted for the greater part of the genetic diversity previously reported from larger areas of the north-eastern United States and central and northern Europe. Although *ospC* alleles of both species had higher nucleotide diversity than other loci, the *ospC* locus showed evidence of intragenic recombination and was unsuitable for phylogenetic inference. In contrast, there was no detectable recombination at the IGS locus of *B. burgdorferi*. Moreover, beyond the signature nucleotides that specified 10 IGS genotypes, there were additional nucleotide polymorphisms that defined a total of 24 subtypes. Maximum-likelihood and parsimony cladograms of *B. burgdorferi* aligned IGS sequences revealed the subtype sequences to be terminal branches of clades, and the existence of at least three monophyletic lineages within *B. burgdorferi*. It is concluded that *B. burgdorferi* and *B. afzelii* have greater genetic diversity than had previously been estimated, and that the IGS locus alone is sufficient for strain typing and phylogenetic studies.

†These two authors contributed equally to the study.

Abbreviations: IGS, intergenic spacer; LB, Lyme borreliosis; MLST, multilocus sequence typing.

The GenBank accession numbers for the sequences reported in this paper are: AY275189–AY275212 for *B. burgdorferi rrs–rrlA* IGS types 1–9 and nt10, as well as subtypes of each IGS genotype; AY363692–AY363702 for *B. afzelii rrs–rrlA* IGS types 1–9, and nt10 and nt11; AY275213–AY275225 for *B. burgdorferi ospC* types 1–9 and nt10–nt13; and AY363710–AY363721 for *B. afzelii ospC* types 1–6, 7A, 7B, 8, 9, nt10 and nt11.

Tables showing pairwise nucleotide and amino acid distances between *ospC* alleles and OspC proteins of both *B. burgdorferi* and *B. afzelii* are available as supplementary data with the online version of this paper at http://mic.sgmjournals.org.

INTRODUCTION

The tick-transmitted spirochaete *Borrelia burgdorferi* is the cause of Lyme borreliosis (LB) in North America; *Borrelia afzelii* and *Borrelia garinii*, as well as *B. burgdorferi*, are LB agents in Europe (Hengge *et al.*, 2003; Steere, 2001). In Europe, most cases of erythema migrans, the usual skin manifestation of LB, are caused by *B. afzelii*, while *B. garinii* is more commonly associated with involvement of the nervous system (Busch *et al.*, 1996a, b; Ornstein *et al.*, 2001). On both continents, LB spirochaetes cycle in nature between certain species of *Ixodes* ticks, such as *Ixodes scapularis* in North America and *Ixodes ricinus* in Europe, and warm-blooded vertebrate reservoir hosts (Anderson & Magnarelli, 1992; Gern *et al.*, 1998). A variety of rodents and ground-feeding birds are hosts for larval and nymphal ticks, and are competent as infection reservoirs (Kurtenbach *et al.*, 2002).

Given the wide geographic distributions of the LB agents throughout temperate regions across the northern hemisphere, and the large number of possible vertebrate reservoirs, one might expect that each species comprises several strains, that is, distinct lineages. On the other hand, LB is an emerging infection in developed countries: the spread of endemic disease is associated with reforestation and increasing deer herds in areas formerly devoted to agriculture or industry (Barbour & Fish, 1993). Under these circumstances, an evolutionary bottleneck several decades ago in the north-eastern United States and in Europe, at the height of land usage for agriculture and industry in these areas, might have produced a population structure that featured only a handful of strains within a species.

In fact, studies to date of strain diversity indicate that each *Borrelia* species comprises a variety of strains under what appears to be balancing selection (Dykhuizen *et al.*, 1993; Mathiesen *et al.*, 1997; Qiu *et al.*, 1997). The presence of multiple strains has been suggested by studies of different genetic loci, both chromosomal and plasmid. Loci that have been studied include the chromosomal intergenic spacer (IGS) between the single 16S (*rrs*) rRNA and the first of two 23S (*rrlA*) rRNA genes (Liveris *et al.*, 1995); the spacer between a 5S (*rrf*) and the second 23S (*rrlB*) rRNA gene (Lee *et al.*, 2000; Postic *et al.*, 1994); *ospA*, a plasmid-borne gene for an outer membrane lipoprotein (Barbour *et al.*, 1985; Norris *et al.*, 1999; Qiu *et al.*, 1997; Wilske *et al.*, 1993); *ospC*, a plasmid-borne gene for another outer membrane lipoprotein (Lagal *et al.*, 2003; Livey *et al.*, 1995; Theisen *et al.*, 1995; Wang *et al.*, 1999; Wilske *et al.*, 1995); *p66*, a chromosomal gene for an integral outer-membrane protein (Bunikis *et al.*, 1998; Norris *et al.*, 1997); and variable-number tandem repeat (VNTR) loci on both the chromosome and the plasmids (Farlow *et al.*, 2002).

Most past studies of strain diversity have depended on a passive or retrospective collection of isolates, rather than a prospective study with random sampling of a pre-defined population. Moreover, with the exception of the studies of Dykhuizen and colleagues (Qiu *et al.*, 2002) and Farlow *et al.* (2002), a limitation of the various reports of typing systems has been a lack of cross-referencing of loci and alleles between strains that were examined together. To remedy this partly, we further developed strain typing of *B. burgdorferi* and other LB species by sequencing multiple genetic loci. The future goal is to apply this strain-typing scheme for phylogenetic inference, and as an epidemiologic tool. A different version of the multiple-locus approach has been used in studies of a variety of other bacterial pathogens (Feavers *et al.*, 1999; Maiden *et al.*, 1998).

The present study had the following aims: (i) to estimate the strain variety and diversity of two LB species, *B. burgdorferi* in the United States and *B. afzelii* in Europe, by sequencing four different loci; (ii) to gauge the traces of recombination in selected loci; and (iii) to assess whether a single locus was as informative as several loci for strain differentiation. We found that each species comprised at least nine clonal strains, that small geographic areas well represented the strain diversity to be found in a much larger area, and that a partial sequence of the *rrs–rrlA* IGS region showed no evidence of recombination and is likely sufficient for differentiation of strains of *B. burgdorferi* and *B. afzelii*.

METHODS

Field and clinical specimens. Nymphal ticks were collected in mixed hardwood forests in two locations where LB is endemic: southern Connecticut in the United States and Blekinge County in southern Sweden (Berglund *et al.*, 1995; CDC, 1997). In Connecticut, *I. scapularis* nymphs were collected during late May to August 1999 by drag sampling, as described by Falco & Fish (1992). Collections were conducted on three 2·4 ha sites, located 100 m to 1·5 km apart, on private land of a water company (Tsao, 2000). In Sweden, the forested study area comprised six neighbouring sites, each divided into two squares of 1250 m^2. *I. ricinus* nymphs were collected at the end of September 2001 from clothes of 10 adult volunteers after each has been exposed for about 30 min to the host-seeking ticks. Another source of isolates in Sweden consisted of randomly selected skin biopsies, obtained under informed consent for the purposes of diagnosis of skin rashes suspected to be erythema migrans, a manifestation of LB, from residents of Blekinge County attending a single clinic as patients during 2001 and 2002. Biopsies were placed in BSK II medium (Barbour, 1984), and were then subjected to DNA extraction (see below).

Additional isolates. The following reference isolates of *B. burgdorferi* were cloned by limiting dilution in BSK II medium or as isolated colonies in BSK II solid medium, previously or in the present study: B31 (ATCC 35210), from a tick collected in 1981 in New York (Barbour *et al.*, 1983); N40, from a tick collected in 1986 in New York (Barthold *et al.*, 1988); and HB19 (formerly 245) and 297, from the blood and cerebrospinal fluid, respectively, of LB patients in Connecticut in 1982 (Barbour & Schrumpf, 1986; Steere *et al.*, 1983). Other *B. burgdorferi* isolates included in the study were BVe, cultured in 1985 from a veery, a passerine bird, in Connecticut (Anderson *et al.*, 1986; Barbour *et al.*, 1985), and PAd, an isolate from the skin of a patient with erythema migrans in New York in 1984 (Berger *et al.*, 1985). Two *B. afzelii* reference strains were employed: ACAI, isolated from the skin of a patient in Sweden in 1984 (Åsbrink & Hovmark, 1985), and PKo, isolated from the skin

of a patient in Germany in 1985 (Wilske *et al.*, 1986). *Borrelia* strains were grown in BSK II medium and harvested as described previously (Barbour, 1984).

DNA extraction, PCR and sequencing. Ticks and skin biopsies were homogenized in liquid nitrogen, and total DNA was extracted using the DNeasy Tissue Kit (Qiagen), as described by Beati & Keirans (2001). DNA was eluted with 50 µl deionized sterile water and stored at −20 °C. Cultivated spirochaetes were harvested by centrifugation for 15 min at 5000 ***g***, and then lysed by boiling for 30 min. DNA extracts from ticks and biopsies were initially screened by *Borrelia* genus-specific PCR that targeted the *flaB* gene, as previously described (Barbour *et al.*, 1996). With the exception of *ospA*, DNA fragments of the loci of interest were amplified by a nested PCR procedure, comprising 35 cycles for the first reaction and 40 cycles for the second reaction. The forward (F), forward-nested (Fn), reverse (R) and reverse-nested (Rn) primers and reaction conditions used are given below.

(1) *rrs–rrlA* IGS: F, 5′-GTATGTTTAGTGAGGGGGGTG-3′ (position 2306–2326 of U03396); R, 5′-GGATCATAGCTCAGGTGGTTAG-3′ (3334–3313); Fn, 5′-AGGGGGGTGAAGTCGTAACAAG-3′ (2318–2339); Rn, 5′-GTCTGATAAACCTGAGGTCGGA-3′ (3305–3284); at 94 °C for 30 s, 56 °C for the first reaction and 60 °C for the second reaction for 30 s, and 74 °C for 60 s.

(2) *p66*: F, 5′-GATTTTTCTATATTTGGACACAT-3′ (positions 1211–1233 of X87725); R, 5′-TGTAAATCTTATTAGTTTTTCAAG-3′ (1966–1943); Fn, 5′-CAAAAAAGAAACACCCTCAGATCC-3′ (1252–1275); Rn, 5′-CCTGTTTTTAAATAAATTTTTGTAGCATC-3′ (1935–1907); at 94 °C for 60 s, 50 °C for 120 s, and 74 °C for 120 s.

(3) *ospC*: F, 5′-ATGAAAAAGAATACATTAAGTGC-3′ (positions 306–328 of U01894); R, 5′-ATTAATCTTATAATATTGATTTTAATTAAGG-3′ (963–933); Fn, 5′-TATTAATGACTTTATTTTTATTTATATCT-3′ (331–359); Rn, 5′-TTGATTTTAATTAAGGTTTTTTTGG-3′ (948–924); at 94 °C for 30 s, 52 °C for 30 s, and 74 °C for 30 s.

The *ospA* sequences were amplified by a single PCR reaction with the following primers and conditions: F, 5′-TATTTATTGGGAATAGGTC-3′ (positions 160–178 of X14407), and R, 5′-GACTCAGCACCTTTTTG-3′ (1049-1033), at 94 °C for 30 s, 51 °C for 60 s, and 72 °C for 120 s. The DNA extractions, PCR reaction preparations, and analysis of the products were carried out in three separate laboratories. To monitor for contamination, negative controls were included in the DNA-extraction and PCR procedures. PCR products were cloned in pCR2.1-TOPO vector (Invitrogen) before sequencing, or were sequenced directly, as described by Bunikis *et al.* (1998). Amplicons or plasmid inserts were sequenced in both directions on a CEQ 8000 automated capillary sequencer (Beckman Coulter). Primer target sites were excluded from sequence analysis.

Sequences. Genbank accession nos for sequences described in this paper are given in the footnote above. The *rrs–rrlA* IGS sequences of *B. burgdorferi* were compared with the sequences reported by Wormser *et al.* (1999), and the *ospC* sequences of *B. burgdorferi* and *B. afzelii* strains were compared with sequences reported by Livey *et al.* (1995), Theisen *et al.* (1995), Wang *et al.* (1999) and Wilske *et al.* (1995).

Sequence analysis. Sequences were initially aligned using the CLUSTAL X algorithm (Thompson *et al.*, 1994), and then manually using MacClade 4.04 software (Maddison & Maddison, 2002). Positions with at least two different characters in at least two sequences each were considered polymorphic, and included in the analyses. Singletons, i.e. variant nucleotides found in only one sequence, were ignored. Descriptive statistics and a linkage disequilibrium test (Hill & Roberts, 1968) of the aligned sequences were carried out with version 3.5 of the DnaSP suite of programs (Rozas & Rozas, 1999). Pairwise distances were evaluated by using PAUP* 4.0b10 (Swofford, 2001). GENECONV version 1.81 (www.math.wustl.edu/∼sawyer/mbprogs) was used to perform Sawyer's test for evidence of gene conversion; it examines the null hypothesis that nucleotide substitutions observed in a set of aligned sequences are randomly distributed (Sawyer, 1989). The implementation of Jolley *et al.* (2001) in their START suite of algorithms (http://pubmlst.org/software/analysis/start/) of the maximum chi-squared test of Maynard Smith (1992) was used to identify possible recombination events between pairs of alleles; the significance level (*P* value) for each pairwise analysis was the proportion of 1000 permutations that had maximum chi-squared values greater than or equal to the observed chi-squared value. Phylogenetic analysis was performed on aligned sequences, with or without modification or character weight change, using neighbour-joining, parsimony or maximum-likelihood routines of PAUP*. The model of evolution that most likely accounts for the sequences in the alignment was identified using the program MODELTEST (Posada & Crandall, 1998). Percentage support values for clades were obtained from 1000 bootstrap iterations.

RESULTS

Experimental design and analysis

We first examined, by PCR, cloned populations of reference isolates, other culture isolates of *B. burgdorferi*, and representative samples of ticks from a field site. For the study of *B. burgdorferi*, we analysed four loci: *ospA*, *ospC*, *p66* and IGS [the sequence between the 16S (*rrs*) and a 23S (*rrlA*) rRNA gene]. We then applied the findings on *B. burgdorferi* to study the diversity of genotypes among a similar collection of *B. afzelii* isolates from Europe and a specific field site in southern Sweden. *B. afzelii* was selected because it was the majority species in nymphal ticks and biopsies from patients with LB in this region of Sweden (U. Garpmo and others, unpublished results). For both studies, total DNA extracts from ticks were first screened by PCR with genus-specific *flaB* gene primers, which not only identified extracts containing *B. burgdorferi* or *B. afzelii* DNA, but also a non-LB *Borrelia* species related to *B. miyamotoi* (Scoles *et al.*, 2001). The population structure of this other species, *B. burgdorferi* at the Connecticut site, and *B. afzelii* in Sweden, will be reported subsequently (J. Tsao and others, unpublished results; U. Garpmo and others, unpublished results).

PCR amplicons were either sequenced directly or cloned into a plasmid vector. Cloning was carried out on selected samples when direct sequencing indicated that two or more alleles of a locus were present. Two or more clones were sequenced to resolve possible sequencing errors, or when a new polymorphic position was suspected. The number of loci determined for each sample depended on the amount of remaining DNA. Review of the minority of samples for which there was incomplete characterization of loci showed that the loci that were determined had a similar distribution of alleles to that of the more completely characterized majority of samples.

The criterion for designating a distinct type, e.g. type 1, of *B. burgdorferi* or *B. afzelii* was the presence of a unique

complement of alleles or sequence variants at each of the loci examined: four loci for *B. burgdorferi* and two loci for *B. afzelii*. Strain B31, the type strain, was arbitrarily assigned to type 1 of *B. burgdorferi*. If a new and unique sequence variant was found at one locus, but not at all other loci of the same sample, then that allele was provisionally assigned the next number in sequence, and also 'nt' (for non-typable) as a prefix, until corresponding variants at other loci, if present, could be identified in future studies. The designation 'subtype' was applied when further polymorphisms distinguished alleles at one locus in the absence of additional polymorphisms at the other loci.

Sequence diversity of *B. burgdorferi* at four loci

DNA was extracted from cultivated isolates of B31, 297, N40, HB19, BVe and PAd, which were tick and vertebrate isolates collected in the 1980s in Connecticut and adjacent New York. Of 1784 *I. scapularis* nymphs collected at the Connecticut field site, 624 (35 %) were positive by *flaB* PCR for *B. burgdorferi*. A random sample of 71 *flaB*-positive extracts was then subjected to amplification and sequencing of 4 loci (61 extracts), 3 loci (9 extracts), or 2 loci (1 extract).

Fig. 1 shows the regions of the *rrs–rrlA* IGS (a) and the *p66* gene (b) that were amplified. The outer forward primer for the partial IGS locus was at the 3′ end of the *rrs* gene, and the outer reverse primer was in the coding sequence for the *ileT* tRNA gene in the spacer. Included in the amplicon was the *alaT* gene for tRNA, and, for the IGS, 805 nt (26 %) of a total of 3052 nt, by *B. burgdorferi* B31 coordinates, were analysed. The analysed fragment of the *p66* gene was 631 nt (34 %) of the 1857 nt total length, and included a sequence encoding a surface-exposed loop, complete and partial predicted loops, and flanking transmembrane segments (Bunikis *et al.*, 1998). Using primers representing the sequences of the conserved 5′ and 3′ ends, we amplified and sequenced 794 nt (97 %) of the 822 nt total length of *ospA*, and 566 nt (89 %) of the 633 nt total length of *ospC*. Table 1 provides descriptive statistics on the alleles of the four loci of *B. burgdorferi*.

Among DNA sequences from the 6 reference strains and 62 extracts of infected ticks, there were 24 sequence variants of the IGS region that ranged from 805 to 812 nt in length (Table 1). For non-gapped positions, the mean nucleotide diversity per position (π) was 0·03; there were no substitutions in the *alaT* coding regions. Table 2 shows the signature nucleotides for 10 major genotypes for the IGS locus; 7 polymorphic positions and 1 indel over a 250 bp length were sufficient to discriminate between the genotypes. Nine of these genotypes corresponded to the unique alleles at other loci (Table 3); the tenth allele was provisionally designated 'nt10'.

Additional nucleotide polymorphisms in the IGS locus defined 21 subtypes within 7 of the 10 genotypes (Table 4). Fig. 2 shows an unrooted maximum-likelihood phylogram of the 24 IGS variants and their classification by the criteria of Liveris *et al.* (1995). IGS types 1 and 3, and their subtypes, would be considered group I by Liveris *et al.* (1995), type 2 subtypes correspond to group II, and the other 7 types and their subtypes would be classified as group III.

Sequencing of *ospC* genes from 6 reference strains and 61 tick samples revealed 13 sequence variants, ranging from 561 to 573 nt in length, and producing 585 aligned characters (Table 1). All but allele 9, which was found in reference strain N40, and allele 13, which was found in the PAd isolate, were identified in at least one tick extract. The nucleotide diversity per position (π) for the *ospC* alignment was 0·19, sixfold higher than π for the IGS. Pairwise nucleotide

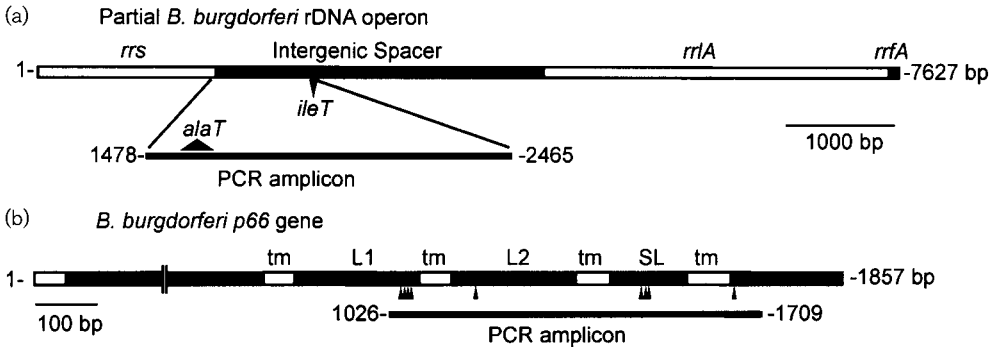

Fig. 1. Physical maps of the partial rRNA operon (a) and the full-length *p66* gene (b) of *B. burgdorferi*. In both panels, the nucleotide positions for the 5′ and 3′ ends of the PCR amplicons are shown; numbering for both loci follows *B. burgdorferi strain* B31 coordinates. In the rRNA operon (a), the *rrs–rrlA* intergenic spacer (IGS) separates the *rrs* (16S) gene from *rrlA–rrfA*, the first of two 23S–5S arrays on the chromosome. The forward primer for the IGS amplicon is at the 3′ end of the *rrs* and the reverse primer is in the *ileT* gene. Included in the IGS amplicon is the *alaT* gene. The map of the *p66* gene (b) indicates the locations of transmembrane (tm) segments separating predicted loops (L1 and L2), as well as the empirically confirmed surface-exposed loop (SL) of the protein. Arrowheads indicate the positions of non-synonymous substitutions and an indel in the amplified fragment of *p66*.

Table 1. Descriptive statistics and Sawyer's test for recombination of selected loci of two *Borrelia* species

π, Mean nucleotide diversity at each aligned position. SD, Number of standard deviations above the mean of 10 000 permutations using GENECONV (Sawyer, 1989; available at http://www.math.wustl.edu/~sawyer/mbprogs/). *P* value, simulated *P* value based on 10 000 permutations with Bonferroni correction for multiple samples. Significant fragments, number of inner fragments with Bonferroni-corrected Karlin–Altschul *P* values of $<0{\cdot}05$. IGS, *rrs–rrlA* intergenic spacer.

Species	Locus	No. of samples	No. of alleles	Aligned characters				Sawyer's test			
				Base pairs	No. gapped	Polymorphisms (%)	π	Max. score	SD	*P* value	Significant fragments
B. burgdorferi	IGS	68	24	812	11	60 (7·4)	0·025	2·7	−0·34	0·57	0
	ospC	67	13	585	30	267 (45·6)	0·185	12·9	7·38	$<10^{-3}$	4
	ospA	71	11	794	0	11 (1·4)	0·004				
	p66	71	12	634	3	18 (2·8)	0·008				
B. afzelii	IGS	107	11	400	0	17 (4·3)	0·013	3·1	2·59	0·02	0
	ospC	77	12	582	30	196 (33·7)	0·136	20·4	14·23	$<10^{-4}$	10

sequence dissimilarity ranged from 14·2 % to 22·0 %, corresponding to amino acid divergences of 21·4 % to 33·9 % (Supplementary data Table S1 at http://mic.sgmjournals.org). A unique *ospC* genotype was defined as a nucleotide sequence that differed by at least 9 % from any other *ospC* sequence. There were no further polymorphisms within each of the 13 variant genotypes among 67 samples of *B. burgdorferi*. Of the *ospC* alleles, 9 appear to be linked to 9 unique IGS genotypes (Table 3). The remainder of the *ospC* alleles were provisionally designated nt10 to nt13.

In comparison to the *ospC* and IGS loci, the plasmid-borne *ospA* and chromosome-borne *p66* genes had lower nucleotide diversities, at 0·004 and 0·01, respectively. Nevertheless, there were as many unique alleles at these loci as at the IGS and *ospC* loci (Table 1). The 11 *ospA* sequences of 794 nt each were defined by 11 polymorphic sites (Table 5). There were five non-synonymous substitutions in the coding region for the processed OspA lipoprotein, and these defined six OspA peptide variants. Nine of the *ospA* alleles were linked to IGS and *ospC* alleles (Table 3).

The 12 sequence variants of a portion of the *p66* gene were either 631 or 634 nt, and were distinguishable at 18 positions (Table 6). Three (38 %) of the eight non-synonymous substitutions were concentrated in the 132 bp region that encodes the known surface loop of P66 (Fig. 1b). Nine of the *p66* alleles were linked to unique IGS, *ospC* and *ospA* alleles (Table 3).

Genetic diversity of *B. afzelii*

For the subsequent study of isolates from Sweden, we focused on the IGS and *ospC* loci. There are several *ospC*

Table 2. Minimal matrix for differentiation of *rrs–rrlA* intergenic spacer (IGS) genotype of *B. burgdorferi*

Reference strains were B31, 297, Bve, HB19 and N40; the remainder were PCR samples from infected *I. scapularis* ticks from the Connecticut field site. Alignment position refers to *B. burgdorferi* B31 coordinates (accession no. U03396). Typing refers to the intergenic spacer typing system of Wormser *et al.* (1999). Δ, 7 bp insertion after position 124. Signature character(s) for each type are underlined. nt, Provisional designation for IGS genotype without identified linkage to allele of other loci of the species.

IS genotype	Sample/ strain	Alignment position								Typing
		62	77	Δ	137	287	289	294	309	
1	2-39/B31	A	C	No	T	T	A	G	G	I
2	1-65/297	G	C	No	C	C	A	A	G	II
3	1-24	A	C	No	G	T	A	G	G	I
4	1-17	A	C	Yes	C	C	A	A	G	III
5	1-66	G	C	Yes	C	T	A	G	G	III
6	1-43/Bve	A	C	Yes	C	C	A	G	A	III
7	4-55/HB19	A	C	Yes	C	C	A	G	G	III
8	1-48	G	T	Yes	C	T	A	G	G	III
9	-/N40	A	C	Yes	C	C	G	G	A	III
nt10	5-6	A	C	Yes	C	T	A	G	G	III

Table 3. Linkage among alleles at different loci for typable strains of *B. burgdorferi* and *B. afzelii* and comparison with other typing systems

Sample/isolate, tick or biopsy sample identical to clonal reference isolate or other culture isolate. Positions 220–562 of the *ospA* gene were amplified for typing by SUNY SB (Guttman *et al.*, 1996; see Table 5). *ospC* group from Wang *et al.* (1999) and Seinost *et al.* (1999); IP *ospC* group from Lagal *et al.* (2003); NYMC IGS RFLP group from Liveris *et al.* (1995).

Multi-locus genotype	Sample/ isolate	Allele at each genetic locus				SUNY SB		IP *ospC* group	NYMC IGS RFLP group
		IGS	*ospC*	*ospA*	*p66*	*ospA* class	*ospC* group		
B. burgdorferi									
1	2-39/B31	1	1	1	1	1	A	B4	I
2	1-65/297	2	2	2	2	3	K	B1	II
3	1-24	3	3	3	3	4	B	B2	I
4	1-17	4	4	4	4	2	N		III
5	1-66	5	5	5	5	1	D		III
6	-/Bve	6	6	6	6	3	M		III
7	4-55/HB19	7	7	7	7	4	I	B3	III
8	1-48	8	8	8	8	4	U		III
9	-/N40	9	9	9	9	2	E		III
B. afzelii									
1	191/Pko	1	1					A2	
2	476	2	2						
3	055	3	3					A3	
4	020	4	4					A8	
5	284	5	5						
6	005	6	6						
7	767	7	7					A6	
8	040	8	8					A7	
9	474/ACAI	9	9					A1	

sequences of *B. afzelii* in nucleotide databases, and a comparison of these sequences with known *ospC* sequences of *B. burgdorferi* indicated that *B. burgdorferi ospC* primers would be suitable for PCR of *B. afzelii ospC* as well. Much less was known of the sequence between the *rrs* and *rrl* genes of *B. afzelii*, but a pilot study showed that the *rrs* and *ileT* sequences were sufficiently conserved between *B. burgdorferi* and *B. afzelii* for the same primers to amplify the IGS of both species. The analysed region of the *B. afzelii* spacer was 400 nt, instead of the 812 nt for *B. burgdorferi*, and there were no gaps in the aligned *B. afzelii* sequences (Table 1). Unlike the *B. burgdorferi* genome (Fraser *et al.*, 1997), an *alaT* tRNA gene or pseudogene was not detected in this part of the *B. afzelii* genome, and the nucleotide diversity of the *B. afzelii* IGS region was also lower than that of *B. burgdorferi* (Table 1).

Among the 107 *B. afzelii* sequences, of which 73 were from infected *I. ricinus* nymphs and 34 were from positive skin biopsies, there were 11 *rrs–rrlA* IGS types, which were defined by their signature nucleotides at 17 polymorphic positions (Tables 1 and 7). Unlike the finding with *B. burgdorferi* IGS genotypes, we observed no additional polymorphisms or subtypes beyond the 11 genotypes among the 107 sequences (Fig. 2).

The amount of *B. afzelii* strain diversity observed at the IGS locus was exceeded by the diversity at the *ospC* locus. There were 11 *ospC* variants, which ranged in pairwise nucleotide dissimilarity from 9·4 % to 19·3 % and deduced amino acid dissimilarity from 10·1 % to 30·1 %, among 77 *B. afzelii* samples from 47 nymphs and 30 positive skin biopsies (Table 1 and Supplementary Table S2 at http://mic.sgmjournals.org). Two variants varied by only 0·5 % in nucleotide sequence and only 1·0 % in deduced amino acid sequence; these variants were provisionally designated subtypes A and B of genotype 7. Of the unique *ospC* types, 9 corresponded to unique *rrs–rrlA* IGS types (Table 3); for 3 alleles, 7B, nt10 and nt11, the linkage or association with a unique *ospC* allele has yet to be identified (Supplementary Table S2 at http://mic.sgmjournals.org).

Linkage of alleles of *B. burgdorferi* and *B. afzelii*

Table 3 summarizes the relationships between the different genotypes for reference isolates, other isolates, and PCR samples that provided the full complement of loci for analysis. Two loci, IGS and *ospC*, were sequenced for the study of *B. afzelii*, and nine distinct linkage groups of unique IGS and *ospC* sequences were identified. In the study of *B. burgdorferi*, 9 of 12 *ospC*, 9 of 11 *ospA*, and 9 of 12 *p66* alleles each corresponded to a unique *rrs–rrlA* IGS type, thus defining 9 multi-locus genotypes of *B.*

Table 4. Polymorphic positions of subtypes of genotypes 1–4 and 6–8 of the *rrs–rrlA* intergenic spacer (IGS) of *B. burgdorferi*

The *rrs–rrlA* IGS position for subtypes is based on *B. burgdorferi* B31 coordinates (accession no. U03396). No subtypes were found for *rrs–rrlA* IGS genotypes 5, 9 and nt10.

Subtype	Sample/ isolate	*rrs–rrlA* IGS position for subtypes																
		321	325	384	408	445	458	473	505	619	628	639	665	692	701	734	796	847
1A	2-39/B31	A																
1B	4-43	G																
2A	2-24			T				A		T				C	C			
2B	-/297			T				C		C				A	C			
2C	14-25			T				A		T				A	T			
2D	4-28			C				A		T				C	C			
3A	1-24					T										T		T
3B	5-17					T										T		C
3C	4-62					T										C		T
3D	5-2					C										T		T
4A	1-17							A	G									
4B	5-13							C	A									
6A	5-20/Bve		A				C				G		C					
6B	2-43a		G				C				T		T					
6C	1-43		A				T				G		C					
7A	-/HB19															T		
7B	4-55															C		
8A	1-48				T		T					T					G	T
8B	10-33				T		C					G					A	C
8C	1-73				T		T					G					A	C
8D	2-3				C		T					G					G	T

burgdorferi. Five of the reference isolates represented different multi-locus types: B31 (type 1), 297 (type 2), BVe (type 6), HB19 (type 7) and N40 (type 9). The sixth reference isolate, PAd, an uncloned, low-passage isolate from a skin biopsy, had a novel *ospC* allele, nt13, but the other loci found in the DNA extract were that of another genotype (IGS type 4, *ospA* type 3, and *p66* type 4), suggesting a mixture of strains in the infected skin (Seinost *et al.*, 1999b). There was also evidence of more than one strain in individual ticks: more than one allele, or discordant alleles for different loci, were identified in 34 (48 %) of the 71 *I. scapularis* ticks infected with *B. burgdorferi*. This high prevalence of mixed infections is consistent with the findings of Guttman *et al.* (1996) and Qiu *et al.* (2002) for *I. scapularis* nymphs in the north-eastern United States. Among samples suspected to contain a mixture of alleles, unique allelic types of *ospC* (nt10–nt13), *ospA* (nt10 and nt11), and *p66* (nt10–nt12) were found, without identified linkage to comparable unique alleles or genotypes at other loci.

Comparison with other sequences and typing methods

The *B. burgdorferi* and *B. afzelii ospC* sequences determined for this study were compared with database sequences of other isolates of these two species to assess the representativeness of our sampling. Each set of sequences was aligned, and unrooted, neighbour-joining distance phylograms with $\geqslant$80 % bootstrap support for each node were determined (Fig. 3). Fig. 3(c) shows the distribution of polymorphic sites over the lengths of *ospC* alleles of *B. burgdorferi* and *B. afzelii*.

The *B. burgdorferi* sequences from the literature included the set of 'OC' alleles reported by Wang *et al.* (1999) from their study of tick isolates from different sites on Long Island, New York. Fig. 3(a) also includes two alleles of *B. burgdorferi* from isolates from California (Son188 and CA-11.2A), and two alleles from isolates from Germany (PKa and ZS7). Of the total of 15 *ospC* genotypes included in the figure, 12 (types 1–12) were found within the 7 ha collecting grid of the Connecticut site. The three exceptions were type 13, represented by PAd and Son188, and types represented by OC11 and OC3 sequences.

A similar phenomenon was observed with the *B. afzelii ospC* alleles. Of the 16 *ospC* types found among isolates from various locations in Europe, including Slovenia, Austria, Switzerland, Germany, France and Denmark, 8 were identified in ticks collected from a 15 000 m^2 area of forest in southern Sweden.

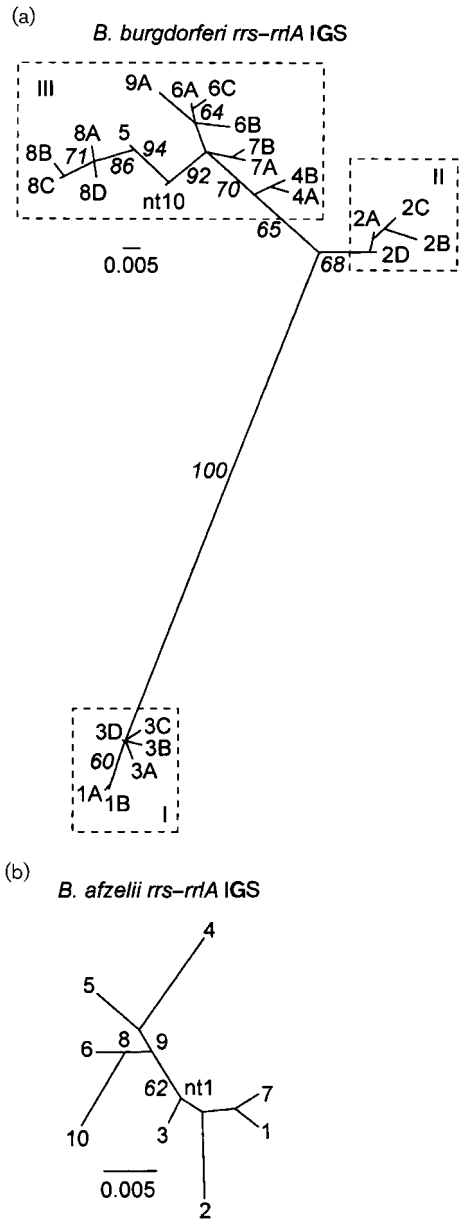

Fig. 2. Unrooted phylograms of aligned partial sequences of the *rrs–rrlA* intergenic spacer (IGS) loci of *B. burgdorferi* (a) and *B. afzelii* (b) strains. Numbers indicate the IGS genotype; subtypes are indicated by letters. Heuristic and exhaustive search options of the maximum-likelihood routine were used for the *B. burgdorferi* and *B. afzelii* trees, respectively. Maximum-likelihood settings were estimated using MODELTEST (Posada & Crandall, 1998). The best-fit evolutionary model was General Time Reversible with gamma rate variation (shape parameter 0·0154) for the *B. burgdorferi* dataset, and Hasegawa–Kishino–Yano with no invariable sites for the *B. afzelii* dataset. Characters were equally weighted, except for the G79T (weight = 5) and A788C transversions (weight = 2), which were introduced in the *B. burgdorferi* IGS alignment to substitute for a 7 bp indel at position 79 and a 4 bp indel at position 788, respectively. Support for clades was evaluated by 1000 bootstrap replications, and values above 60 % are indicated in italics along branches. The scale bars indicate the number of substitutions per site. Sequences conforming to *rrs–rrlA* IGS groups I, II and III of *B. burgdorferi* by the criteria of Liveris *et al.* (1995) are demarcated by dashed lines.

Table 3 shows how the nine multi-locus types would likely be classified by other typing systems: three based on single-strand conformation polymorphisms or reverse line-blotting of *ospA* or *ospC* (Lagal *et al.*, 2003; Qiu *et al.*, 1997, 2002), and one based on RFLPs of the *rrs–rrlA* IGS (Liveris *et al.*, 1995). Each of the nine multi-locus types of *B. burgdorferi* corresponded with a different *ospC* group of Wang *et al.* (1999) and Seinost *et al.* (1999a). However, the four categories of *ospA* typing system, and the three categories of *rrs–rrlA* IGS RFLP typing, even if used in combination, could not distinguish between the nine types (Liveris *et al.*, 1995; Qiu *et al.*, 1997). Four multi-locus types of *B. burgdorferi* and the majority of multi-locus types of *B. afzelii* from the present study could be identified with *ospC* groups found by Lagal *et al.* (2003).

Evaluation of recombination

Sawyer's test assesses the likelihood for a set of aligned homologous sequences that the polymorphic fragments arose through recombination rather than mutation (Sawyer, 1989). The test is appropriate for sets of sequences with the level of nucleotide diversity shown by the *ospC* and IGS sequences, but not for the *ospA* and *p66* alleles, which do not have a sufficient number of informative polymorphic sites. As summarized in Table 1, the Sawyer's test results confirmed the extensive recombination of *ospC* alleles of *B. burgdorferi* and *B. afzelii* (Dykhuizen & Baranton, 2001; Livey *et al.*, 1995; Theisen *et al.*, 1995). The maximum chi-squared test was applied in 36 pairwise comparisons to the 459 non-gapped positions of *ospC* alleles 1–9 of *B. burgdorferi* listed in Table 3. The results provided additional evidence of recombination among *ospC* alleles. Using 1000 permutations with the same number of informative sites for each pairwise comparison, we found that the *P* value was $<0{\cdot}01$ for all 36 pairs, and $<0{\cdot}05$ for 33 of the 36 pairs after applying the Bonferroni correction for multiple comparisons.

There was also evidence of recombination of *ospC* alleles between *B. burgdorferi* and *B. afzelii* or, more likely, in a common ancestor. The only characters that distinguished between the set of *B. burgdorferi ospC* alleles and those of *B. afzelii* in an alignment were at positions 20, 25, 28, 43 and 59 at the 5′ end of the gene, which shows the least nucleotide diversity between alleles (Fig. 3c).

On the other hand, we found no evidence of recombination at the IGS locus of *B. burgdorferi*. By Sawyer's test (Table 1) there was one significant inner fragment detected among the less-diverse and shorter IGS sequences of *B. afzelii*, but this was rejected by the more conservative Karlin–Altschul criteria. The parsimony informative sites in the

Table 5. *ospA* alleles of *B. burgdorferi* from the north-eastern United States

ospA polymorphic sites are based on *B. burgdorferi* B31 coordinates (accession no. X14407). Positions with non-synonymous substitutions are underlined. nt, Provisional designation for *ospA* allele without identified linkage to allele or genotype of other loci.

ospA allele	Sample/ isolate	*ospA* polymorphic sites										
		45	117	201	240	348	382	446	465	511	689	817
1	2-39/B31	A	A	T	A	C	G	G	T	C	A	A
2	1-65/297	A	A	T	A	C	G	A	T	C	A	A
3	1-24	A	A	T	A	C	G	A	C	C	A	A
4	1-17	A	A	T	A	C	G	A	T	T	A	A
5	1-66	A	A	T	A	T	G	G	T	C	A	A
6	5-20/Bve	A	A	T	A	C	A	A	T	C	A	A
7	4-55/HB19	A	C	C	A	C	G	A	C	C	A	A
8	1-48	A	C	T	A	C	G	A	T	T	A	C
9	-/N40	A	C	T	A	C	G	A	T	T	A	A
nt10	2-24	A	A	T	A	C	G	A	T	C	G	A
nt11	5-6	G	A	T	G	C	G	A	T	T	A	A

B. burgdorferi IGS sequence alignments were examined for linkage disequilibrium. For the 24 subtype alleles of the *B. burgdorferi* IGS locus, 255 (16·6 %) of 1540 pairwise comparisons of 56 polymorphic sites were significant ($P<0·05$) after Bonferroni correction of Fisher's exact test. The maximum chi-squared test was applied to the 803 non-gapped positions of IGS genotypes 1–9 of *B. burgdorferi* listed in Table 3. Even without Bonferroni correction, the *P* values were $>0·05$ for all 36 pairs and were $>0·10$ for 33 of 36 pairs. Similarly, when applied to the 378 non-gapped nucleotides of IGS genotypes 1–11 of *B. afzelii*, *P* was $>0·10$ for all 55 pairwise comparisons.

Strain phylogeny

Although the major emphasis of the work was on strain diversity and validation of a genotyping protocol, the data were also examined with regard to strain phylogeny. The

Table 6. *p66* alleles of *B. burgdorferi* from the north-eastern United States

p66 polymorphic sites numbered according to *B. burgdorferi* B31 coordinates (accession no. X87725); positions with non-synonymous substitutions are underlined. nt, Provisional designation for *p66* allele without identified linkage to allele or genotype of other loci.

p66 allele	Sample/ isolate	*p66* polymorphic sites																			
		1064	1065	1072	1079	1091	1191	1243	1249	1273	1289	1293	1363	1389	1392	1414	1448	1502	1508	1542	1625
1	2-39/B31	A		G	A	C	C	T	T	T	G	C	G	T	C	C	A	G	G	A	
2	1-65/297	A		G	A	T	C	C	T	C	G	C	T	C	C	C	A	G	G	A	
3	1-24	A		G	A	C	C	T	T	T	A	C	T	C	C	C	A	G	G	A	
4	1-17	A		G	A	C	C	T	T	T	G	C	T	T	T	A	G	G	G	A	
5	1-66	A		G	A	T	C	C	T	C	G	C	T	C	C	C	A	A	G	A	
6	1-43/Bve	A		G	A	T	C	C	T	T	G	C	T	C	C	C	A	G	T	A	
7*	4-55/HB19	A	AAT	G	A	C	T	C	A	T	G	T	T	C	C	C	A	G	G	A	
8	1-48	A		A	A	T	C	C	T	C	G	C	T	C	C	C	A	G	G	A	
9	-/N40	A		G	A	C	C	T	T	T	G	C	T	C	C	C	A	G	G	A	
nt10	4-43	A		G	A	C	C	C	T	T	G	C	T	C	C	C	A	G	G	A	
nt11	5-20	A		G	A	C	C	T	T	T	G	C	T	T	T	A	G	G	T	A	
nt12	11-16	G		G	G	C	C	T	T	T	G	C	T	T	T	A	G	G	G	G	

*3 nt insertion at position 1065 of *p66* allele.

Table 7. *rrs–rrlA* intergenic spacer (IGS) genotypes of *B. afzelii* from southern Sweden

rrs-rrlA IGS first position defined by *B. burgdorferi* B31 *rrs–rrlA* IGS coordinates (accession no. U03396). Genotype 6 found in biopsies, but not in ticks. nt, Provisional designation for IGS genotype of *B. afzelii* without identified linkage to *ospC* allele of the species.

IGS genotype	PCR sample	*rrs–rrlA* IGS position																
		1 8	2 4	4 5	7 6	7 8	9 3	1 3 0	1 4 7	2 1 3	2 1 5	2 5 1	2 5 9	2 6 7	2 8 7	3 2 4	3 4 2	3 5 6
1	191	T	C	A	T	C	G	T	C	T	A	T	T	A	C	T	G	A
2	476	T	C	A	T	C	G	T	T	T	C	T	A	A	C	C	G	G
3	055	C	T	A	T	C	G	T	C	T	C	T	A	A	C	T	G	A
4	020	C	C	G	G	T	A	T	C	T	C	G	T	A	C	T	G	G
5	284	C	C	A	T	C	G	T	C	G	C	G	T	C	C	T	G	G
6	005	C	C	A	T	C	G	A	C	T	C	G	T	A	T	T	G	A
7	767	T	C	G	T	C	G	T	C	T	A	T	A	A	C	T	G	A
8	040	C	C	A	T	C	G	T	C	T	C	G	T	A	T	T	G	A
9	474	C	T	A	T	C	G	T	C	T	C	G	T	A	T	C	A	A
nt10	437	C	C	A	T	C	G	T	C	T	C	T	A	A	C	T	G	A
nt11	025	C	C	A	T	C	G	T	C	T	C	G	T	A	C	T	G	A

ospC alleles were highly polymorphic, with many parsimony informative sites (Table 1), and, moreover, they were linked with almost all of the IGS alleles in B. *burgdorferi* and *B. afzelii* (Table 3). However, the *ospC* locus is not suitable for phylogenetic studies, because of the extensive intra- and inter-species recombination at this locus, as shown in this study (Table 1) and by others (Dykhuizen & Baranton, 2001; Livey *et al.*, 1995; Theisen *et al.*, 1995).

The IGS locus, on the other hand, had lower nucleotide diversity than the *ospC* locus but, at least for *B. burgdorferi*, there was no evidence of intragenic recombination, which could confound attempts to identify monophyletic groups. Fig. 2(a) shows an unrooted maximum-likelihood tree with bootstrap support for nodes for the 24 different IGS listed in Table 4. The figure shows that almost all of the sequences that we designated as subtypes of a particular type were more closely related to each other than they were to a subtype of another type. In other words, they were monophyletic. There was also strong support for a clade of types 1 and 3, a clade of type 2, and a clade of types 5, 8 and nt10. There was less support for the topology of type 4, 6, 7 and 9 genotypes. A maximum-parsimony tree based on *rrs–rrlA* IGS alignment with indels treated as the fifth character showed identical topology of and similar support values for clades (not shown). Phylogenetic analysis of the less polymorphic *B. afzelii* dataset produced a maximum-likelihood tree with some support for two, albeit closely related, clusters of strains (Fig. 2b).

DISCUSSION

Taxonomic and typing studies soon followed the first cultivation of a hitherto unknown group of spirochaetes from LB patients and from the vertebrate reservoirs and tick vectors for the infection. The early surveys of isolates examined protein profiles by PAGE, the binding of monoclonal and polyclonal antibodies in immunofluorescence and immunoblot assays, and RFLPs in Southern blot analyses (Barbour & Schrumpf, 1986; Barbour *et al.*, 1985; Wilske *et al.*, 1986). Some of the trait differences between isolates were eventually shown to be interspecific rather than interstrain features. Other characteristics, such as the presence or absence of the OspC protein, were demonstrated to be phenotypic, and not genotypic, characteristics (Schwan *et al.*, 1995).

Although many of the early typing schemes retain some value in estimating strain diversity and distinguishing between strains, a systematic approach towards understanding the structure, temporal and spatial dynamics, and host adaptation of *Borrelia* populations ultimately requires genetic criteria, preferably nucleotide sequence, to define strains. The present study started with the aim of identifying the most efficacious genetic marker to differentiate between different strains of two LB agents from two different locales with highly endemic disease: *B. burgdorferi* from the north-eastern United States and *B. afzelii* from southern Sweden. This was followed by characterization and comparison of polymorphisms at the candidate loci. The bulk of the samples in the study were derived from field collections of the host-seeking ticks that transmit infection to humans.

We used PCR to sample the strain diversity of *B. burgdorferi* and *B. afzelii* in ticks from the study sites, as well as the diversity of selected reference strains and isolates from the 1980s. In the past, most surveys of genetic heterogeneity have depended on strains isolated first in culture medium. *In vitro* cultivation has the advantage of permitting single-cell cloning by limiting dilution or colony plating. However, strains of *B. burgdorferi* and *B. afzelii* may vary in their

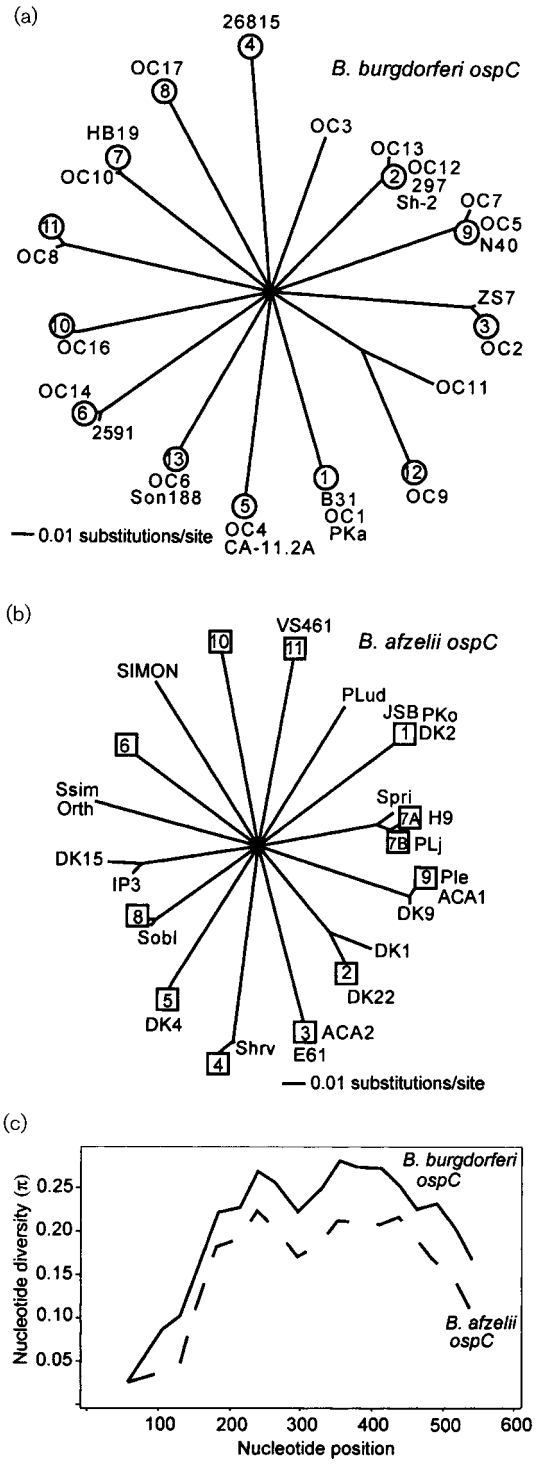

Fig. 3. Unrooted neighbour-joining distance phylograms (with nodes with ⩾80% bootstrap support for 10000 replications) for *ospC* alleles of *B. burgdorferi* (a) and *B. afzelii* (c) strains, together with the distribution of polymorphic sites in *ospC* alleles of both species (c). In (a), numbers in circles refer to *B. burgdorferi ospC* genotypes of Table 3 or Supplementary Table S1 at http://mic.sgmjournals.org; sequences designated 10, 11, 12 and 13 in the figure are genotypes nt10, nt11, nt12 and nt13. Sequences of cloned reference strains B31, 297, N40 and HB19 were determined for this study. Sequences with the 'OC' prefix are from Wang *et al.* (1999). Other sequences were obtained from GenBank (accession nos in parentheses): Sh-2 (AF500203), ZS7 (AF500204), PKa (X69589), CA-11.2A (L25413), Son188 (L81130), 2591 (U01892), 26815 (L42897). In (b), numbers in squares refer to *B. afzelii ospC* genotypes of Table 3 or Supplementary Table S2 at http://mic.sgmjournals.org; sequences designated 10 and 11 in the figure are genotypes nt10 and nt11. Sequences of reference strains PKo and ACAI were determined for this study. Sequences with the 'DK' prefix are from Theisen *et al.* (1995). Sequences of isolates SIMON, E61, Orth, H9 and JSB are from Livey *et al.* (1995). The other sequences were obtained from GenBank (accession nos in parentheses): Shrv (AY150200), Sob1 (AY150201), vs461 (AF416426), PLj7 (X81523), Ssim (AY150202), Spri (AY150204), ACA2 (AY150206), PLud (X83552), IP3 (AF230184), Ple (AB009897). In (c), a sliding window of 100 nt, with a step size of 25 nt, was used for same-length alignments of *B. burgdorferi ospC* alleles 1–9 and nt10–nt13 (Table S1) and *B. afzelii ospC* alleles 1–9 and nt10–nt11 (Table S2). The mean nucleotide diversity (π) along the lengths of the sequences for *B. burgdorferi* and *B. afzelii* is shown.

ability to grow in cell-free culture medium, and consequently uncultivable or poorly growing varieties may not be represented in a collection of cultivated organisms (Liveris *et al.*, 1999; Norris *et al.*, 1997). PCR can bypass the cultivation step, but at the cost of having results confounded by mixtures of strains in the ticks or vertebrate hosts (Guttman *et al.*, 1996; Qiu *et al.*, 2002). A risk, then, of PCR-based sampling of the population is artifactual linkage, for example, between allele *x* at locus A with allele *p* at locus B, when locus A came from one strain and locus B came from another in a mixed infection. Given the evidence that LB *Borrelia* species are rarely subject to horizontal transfer of genes (Ochman *et al.*, 2000), and have a highly clonal population structure (Dykhuizen & Baranton, 2001; Dykhuizen *et al.*, 1993; Wang *et al.*, 1999), we chose to interpret what might be an exception from the normal clonality as the effect of mixed infection instead. This meant that some linkages remained unresolved; such was the case for alleles with the 'nt' prefix. But, overall, most of the results from individual tick extractions were unambiguous, and usually confirmed what had been found with cloned reference strains or the majority of other samples with a particular allele.

We examined four non-paralogous genetic loci in this study, but the analysis was not, strictly speaking, multilocus sequence typing (MLST). As usually practised, MSLT indexes what is assumed to be neutral variation in a set of housekeeping genes of bacterial chromosomes (Maiden *et al.*, 1998). Two of the loci we studied were chromosomal, but one, *p66*, has no known orthologue outside the genus *Borrelia*, and the other, the *rrs–rrl* IGS, is, with the exception of a tRNA gene, a non-coding region. The other two loci

were located on plasmids. Although one of the plasmid loci, *ospA*, by virtue of its limited expression in vertebrates (Schwan *et al.*, 1995), appears not to be under positive selection by innate and adaptive immune systems, the other plasmid locus, *ospC*, which is expressed during mammalian infection, most likely is (Qiu *et al.*, 1997; Wang *et al.*, 1999). While from a certain perspective an analysis of neutral variation in a set of housekeeping genes would be desirable for both phylogenetic and epidemiological purposes, we found this difficult to achieve with *Borrelia* species.

One limitation to the application of MLST, as it is commonly defined, to *B. burgdorferi* has been the comparatively few polymorphisms among conserved genes studied to date in this species. For instance, the *flaB* gene, which encodes a structural protein of the internal flagella of spirochaetes, differed at only 2 nucleotide positions out of 600 nucleotides of the partial sequence between strains B31 (accession no. AB035615), 297 (AB035616) and HB19 (X75200), which represent types 1, 2 and 7, respectively (Table 3). In over 396 nt of another region of *flaB*, there were no differences in sequence between B31 (AF416433) and N40 (AF16447), a representative of type 9. In comparison, B31 differed from N40 at five positions in the IGS loci (Table 2), three positions in the *ospA* alleles (Table 5), and two positions in the *p66* alleles (Table 6). In any case, as further discussed below, the highly clonal population structure of *B. burgdorferi* suggests that for typing and epidemiological purposes, at least, a single locus with several informative sites, but without evidence of recombination, was sufficient.

The population structure of *B. burgdorferi* and *B. afzelii* at the field sites in the USA and Sweden will be considered in subsequent reports, but the findings of the present study can be viewed as confirming the report of Wang *et al.* (1999) that genetic diversity at the local level is representative of the diversity found over a much larger geographic area. The different *ospC* alleles documented in ticks at the small site in Connecticut, and the even smaller site in Sweden, accounted for most of the *ospC* alleles previously found and characterized in the USA and Europe, respectively. Many of the sequences from this and other studies are from reference and other isolates dating from the 1980s. The complement of strains at a given enzootic location seems to represent most, but not necessarily all, of the diversity of *B. burgdorferi* and *B. afzelii* observed over the last two decades. This could be because of the wide dispersal of strains by birds and other mobile hosts. However, we think an equally or more likely explanation is that reservoir host-diversity maintains spirochaete diversity, and the species composition of reservoir hosts is similar on the local and regional scales.

Although most of the major *ospC* groups found in a survey of New York ticks were found in infected ticks at the Connecticut site (Fig. 3), we could not confirm the findings of Wang *et al.* (1999) of further nucleotide polymorphisms among *ospC* alleles within a particular group. Among all the *ospC* sequences from *B. burgdorferi* that we examined, there were none that differed from another allele by less than 9 %. Two *B. afzelii ospC* alleles were an exception: two variants, designated 7A and 7B, were identical, except at their 3′ ends. Whereas *ospC* type 7A was linked to a unique IGS type, 7, a unique linkage has yet to be established for the 7B allele. One possibility is that this represents a lateral gene transfer of most of one *ospC* allele to another strain, as suggested by Jauris-Heipke *et al.* (1995) in their report of some apparently chimeric *ospC* genes of *B. afzelii* strains. As this and other studies have shown, *ospC* alleles are highly divergent in sequence and are suitable for strain identification. However, the star pattern and the long terminal branches of phylograms of *ospC* sequences in Fig. 3 indicate the low value of this locus for inferences about *B. burgdorferi* or *B. afzelii* evolution (Posada & Crandall, 2002).

Neither of the other two loci examined in this study, *p66* and *ospA*, provided added value for phylogenetic inferences either, but this was because of the limited number of parsimony informative sites, rather than evidence of intragenic recombination. Nevertheless, the study demonstrated greater diversity at these two loci within *B. burgdorferi* than had previously been recognized, and this may have implications for studies of LB pathogenesis and immunity. For *p66*, the concentration of non-synonymous substitutions in that part of the sequence that encodes the surface-exposed region of this outer-membrane protein supports the proposal that this protein is under positive selection (Bunikis *et al.*, 1998). P66 is also one of the diagnostic antigens in an immunoblotting assay for LB (Dressler *et al.*, 1993), and it is possible that the antigenic heterogeneity of P66 can affect serodiagnostic test performance, especially if a non-representative strain is chosen as the antigen source.

For the analysis of *ospA* alleles, the sequenced PCR products were longer than those employed in past surveys, and this led to the discovery of several novel *ospA* genotypes in *B. burgdorferi* (Guttman *et al.*, 1996; Qiu *et al.*, 2002). This, in turn, allowed determination of linkages of the distinct *ospA* alleles with alleles at other loci (Table 3). Whether these phenotypic differences in OspA are adaptive is not known. Previous studies of *B. burgdorferi* OspA with monoclonal antibodies showed only a single antigenic type among a limited number of isolates examined (Barbour & Schrumpf, 1986; Wilske *et al.*, 1993). On the other hand, there would be 6 different variants of the mature protein from the 11 different *ospA* alleles identified in this study. We predict that most of the OspA polypeptides would be bound by the neutralizing antibodies elicited by the OspA vaccine for dogs and humans (Ding *et al.*, 2000). A possible exception is the OspA of allele nt10 (Table 5), which had not previously been described and which has an Arg–Lys substitution at position 230, a residue centred in the epitope for these neutralizing antibodies (Ding *et al.*, 2000).

As Schwartz and colleagues suggested (Liveris *et al.*, 1995), there are several advantages of using the *rrs–rrlA* IGS locus both to identify, and to discriminate between, *Borrelia* strains. First, the *rrs* 16S rRNA gene and the *ileT* tRNA gene

are highly conserved in sequence, and so are suitable sites for PCR primers capable of amplifying not only DNA from LB-group species, such as *B. burgdorferi*, as demonstrated here, but also DNA from other *Borrelia* species, such as relapsing fever agents (J. Bunikis and others, unpublished results). With the exception of the *alaT* gene in *B. burgdorferi*, the sequence between *rrs* and *ileT* is highly polymorphic for a chromosomal locus, and, unlike most plasmid-borne genes, the IGS locus is unlikely to be lost from *B. burgdorferi* during propagation. Moreover, the IGS appears not to be under either positive or, excepting *alaT*, purifying selection, and does not show evidence of recombination. In these respects, the IGS differs from the *ospC* locus, which shows the effects of positive selection, presumably by immune systems of different hosts, hastened by recombination after lateral transfer.

The major IGS genotypes we found, among both the reference isolates and the field specimens, were linked to unique complements of three other loci, in the case of *B. burgdorferi*, and one other locus, in the case of *B. afzelii*. Although we detected no further heterogeneity of the *ospA*, *p66*, or the even more diverse *ospC* genes of *B. burgdorferi*, most of the major IGS genotypes of this species could be further distinguished at additional polymorphic nucleotide positions. Those alleles with a unique set of signature nucleotides we termed IGS subtypes. It is not known whether a subtype is operationally a 'strain', in the sense of being a clonal, stable lineage.

One justification for considering *B. burgdorferi* IGS subtype sequences as strain markers is that they are terminal branches of clades, by both maximum-likelihood (Fig. 2) and parsimony criteria. The consensus tree indicates that there are at least three monophyletic groups represented by IGS sequences and their subtypes: genotypes 1 and 3, genotype 2, and genotypes 5, 8 and nt10. By the IGS RFLP typing scheme of Liveris *et al.* (1995), the clade with genotypes 1 and 3 corresponds to group I, the clade with the genotype 2 sequences corresponds to group II, and the genotype 5, 8 and nt10 clade would be part of the less-differentiated group III (Table 3). Group I-type strains have been associated with a higher frequency of disseminated infection in humans and more invasive disease in experimental animals (Wang *et al.*, 2002; Wormser *et al.*, 1999; Lagal *et al.*, 2003; Seinost *et al.*, 1999a). Thus, the phylogeny of LB *Borrelia* species has implications not only for epidemiology and population biology, but also for studies of pathogenesis. Finer discrimination between strains, such as has been demonstrated in the present study, may facilitate identification of the traits that contribute to greater virulence or transmissibility.

ACKNOWLEDGEMENTS

We thank Hossein Mirian and Hany Mattaous for technical assistance, and Ingvar Eliasson for advice and support. This work was funded by grants to J. Bunikis (919558-01 from the Centers for Disease Control and Prevention), to J. Tsao (National Science Foundation), to J. Berglund (Medical faculty of Lund University and the County Council of Blekinge), to D. Fish (The Harold G. and Leila Y. Mathers Charitable Foundation and 58-1265-5023 from the United States Department of Agriculture) and to A. Barbour (AI37248 from the National Institutes of Health). The Southern Sweden *Borrelia* Study Group is acknowledged for help with patient recruitment and tick collection.

REFERENCES

Anderson, J. F. & Magnarelli, L. A. (1992). Epizootiology of Lyme disease and methods of cultivating *Borrelia burgdorferi*. *Ann N Y Acad Sci* **653**, 52–63.

Anderson, J. F., Johnson, R. C., Magnarelli, L. A. & Hyde, F. W. (1986). Involvement of birds in the epidemiology of the Lyme disease agent *Borrelia burgdorferi*. *Infect Immun* **51**, 394–396.

Åsbrink, E. & Hovmark, A. (1985). The spirochetal etiology of erythema chronicum migrans Afzelius. *Acta Pathol Microbiol Immunol Scand Sect B* **93**, 161–163.

Barbour, A. G. (1984). Isolation and cultivation of Lyme disease spirochetes. *Yale J Biol Med* **57**, 521–525.

Barbour, A. G. & Fish, D. (1993). The biological and social phenomenon of Lyme disease. *Science* **260**, 1610–1616.

Barbour, A. G. & Schrumpf, M. E. (1986). Polymorphisms of major surface proteins of *Borrelia burgdorferi*. *Zentbl Bakteriol Mikrobiol Hyg [A]* **263**, 83–91.

Barbour, A. G., Burgdorfer, W., Hayes, S. F., Peter, O. & Aeschlimann, A. (1983). Isolation of a cultivable spirochete from *Ixodes ricinus* ticks of Switzerland. *Curr Microbiol* **8**, 123–126.

Barbour, A. G., Heiland, R. A. & Howe, T. R. (1985). Heterogeneity of major proteins in Lyme disease borreliae: a molecular analysis of North American and European isolates. *J Infect Dis* **152**, 478–484.

Barbour, A. G., Maupin, G. O., Teltow, G. J., Carter, C. J. & Piesman, J. (1996). Identification of an uncultivable *Borrelia* species in the hard tick *Amblyomma americanum*: possible agent of a Lyme disease-like illness. *J Infect Dis* **173**, 403–409.

Barthold, S. W., Moody, K. D., Terwilliger, G. A., Jacoby, R. O. & Steere, A. C. (1988). An animal model for Lyme arthritis. *Ann N Y Acad Sci* **539**, 264–273.

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

Berger, B. W., Kaplan, M. H., Rothenberg, I. R. & Barbour, A. G. (1985). Isolation and characterization of the Lyme disease spirochete from the skin of patients with erythema chronicum migrans. *J Am Acad Dermatol* **13**, 444–449.

Berglund, J., Eitrem, R., Ornstein, K. & 7 other authors (1995). An epidemiologic study of Lyme disease in southern Sweden. *N Engl J Med* **333**, 1319–1327.

Bunikis, J., Luke, C. J., Bunikiene, E., Bergstrom, S. & Barbour, A. G. (1998). A surface-exposed region of a novel outer membrane protein (P66) of *Borrelia* spp. is variable in size and sequence. *J Bacteriol* **180**, 1618–1623.

Busch, U., Hizo-Teufel, C., Boehmer, R., Fingerle, V., Nitschko, H., Wilske, B. & Preac-Mursic, V. (1996a). Three species of *Borrelia burgdorferi* sensu lato (*B. burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii*) identified from cerebrospinal fluid isolates by pulsed-field gel electrophoresis and PCR. *J Clin Microbiol* **34**, 1072–1078.

Busch, U., Hizo-Teufel, C., Bohmer, R., Fingerle, V., Rossler, D., Wilske, B. & Preac-Mursic, V. (1996b). *Borrelia burgdorferi* sensu lato strains isolated from cutaneous Lyme borreliosis biopsies

differentiated by pulsed-field gel electrophoresis. *Scand J Infect Dis* **28**, 583–589.

C. D. C. (1997). Lyme disease – United States, 1996. *Morb Mortal Wkly Rep* **46**, 531–535.

Ding, W., Huang, X., Yang, X., Dunn, J. J., Luft, B. J., Koide, S. & Lawson, C. L. (2000). Structural identification of a key protective B-cell epitope in Lyme disease antigen OspA. *J Mol Biol* **302**, 1153–1164.

Dressler, F., Whalen, J. A., Reinhardt, B. N. & Steere, A. C. (1993). Western blotting in the serodiagnosis of Lyme disease. *J Infect Dis* **167**, 392–400.

Dykhuizen, D. E. & Baranton, G. (2001). The implications of a low rate of horizontal transfer in *Borrelia*. *Trends Microbiol* **9**, 344–350.

Dykhuizen, D. E., Polin, D. S., Dunn, J. J., Wilske, B., Preac-Mursic, V., Dattwyler, R. J. & Luft, B. J. (1993). *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. *Proc Natl Acad Sci U S A* **90**, 10163–10167.

Falco, R. C. & Fish, D. (1992). A comparison of methods for sampling the deer tick, *Ixodes dammini*, in a Lyme disease endemic area. *Exp Appl Acarol* **14**, 165–173.

Farlow, J., Postic, D., Smith, K. L., Jay, Z., Baranton, G. & Keim, P. (2002). Strain typing of *Borrelia burgdorferi*, *Borrelia afzelii*, and *Borrelia garinii* by using multiple-locus variable-number tandem repeat analysis. *J Clin Microbiol* **40**, 4612–4618.

Feavers, I. M., Gray, S. J., Urwin, R., Russell, J. E., Bygraves, J. A., Kaczmarski, E. B. & Maiden, M. C. (1999). Multilocus sequence typing and antigen gene sequencing in the investigation of a meningococcal disease outbreak. *J Clin Microbiol* **37**, 3883–3887.

Fraser, C. M., Casjens, S., Huang, W. M. & 35 other authors (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* **390**, 580–586.

Gern, L., Estrada-Pena, A., Frandsen, F. & 7 other authors (1998). European reservoir hosts of *Borrelia burgdorferi* sensu lato. *Zentbl Bakteriol* **287**, 196–204.

Guttman, D. S., Wang, P. W., Wang, I. N., Bosler, E. M., Luft, B. J. & Dykhuizen, D. E. (1996). Multiple infections of *Ixodes scapularis* ticks by *Borrelia burgdorferi* as revealed by single-strand conformation polymorphism analysis. *J Clin Microbiol* **34**, 652–656.

Hengge, U. R., Tannapfel, A., Tyring, S. K., Erbel, R., Arendt, G. & Ruzicka, T. (2003). Lyme borreliosis. *Lancet Infect Dis* **3**, 489–500.

Hill, W. G. & Roberts, A. (1968). Linkage disequilibrium in finite populations. *Theor Appl Genet* **38**, 226–231.

Jauris-Heipke, S., Liegl, G., Preac-Mursic, V., Rossler, D., Schwab, E., Soutschek, E., Will, G. & Wilske, B. (1995). Molecular analysis of genes encoding outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato: relationship to *ospA* genotype and evidence of lateral gene exchange of ospC. *J Clin Microbiol* **33**, 1860–1866.

Jolley, K. A., Feil, E. J., Chan, M.-S. & Maiden, M. C. J. (2001). Sequence type analysis and recombinational tests (START). *Bioinformatics* **17**, 1230–1231.

Kurtenbach, K., De Michelis, S., Sewell, H. S. & 10 other authors (2002). The key roles of selection and migration in the ecology of Lyme borreliosis. *Int J Med Microbiol* **291 Suppl 33**, 152–154.

Lagal, V., Postic, D., Ruzic-Sabljic, E. & Baranton, G. (2003). Genetic diversity among *Borrelia* strains determined by single-strand conformation polymorphism analysis of the *ospC* gene and its association with invasiveness. *J Clin Microbiol* **41**, 5059–5065.

Lee, S. H., Kim, B. J., Kim, J. H., Park, K. H., Yeo, S. J., Kim, S. J. & Kook, Y. H. (2000). Characterization of *Borrelia burgdorferi* strains isolated from Korea by 16S rDNA sequence analysis and PCR-RFLP analysis of *rrf* (5S)–*rrl* (23S) intergenic spacer amplicons. *Int J Syst Evol Microbiol* **50**, 857–863.

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

Liveris, D., Varde, S., Iyer, R. & 8 other authors (1999). Genetic diversity of *Borrelia burgdorferi* in Lyme disease patients as determined by culture versus direct PCR with clinical specimens. *J Clin Microbiol* **37**, 565–569.

Livey, I., Gibbs, C. P., Schuster, R. & Dorner, F. (1995). Evidence for lateral transfer and recombination in OspC variation in Lyme disease *Borrelia*. *Mol Microbiol* **18**, 257–269.

Maddison, D. R. & Maddison, W. P. (2002). *MacClade 4: Analysis of Phylogeny and Character Evolution*, version 4.04. Sunderland, MA: Sinauer Associates.

Maiden, M. C., Bygraves, J. A., Feil, E. & 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* **95**, 3140–3145.

Mathiesen, D. A., Oliver, J. H., Jr, Kolbert C. P. & 9 other authors (1997). Genetic heterogeneity of *Borrelia burgdorferi* in the United States. *J Infect Dis* **175**, 98–107.

Maynard Smith, J. (1992). Analysing the mosaic structure of genes. *J Mol Evol* **34**, 126–129.

Norris, D. E., Johnson, B. J., Piesman, J., Maupin, G. O., Clark, J. L. & Black, W. C., IV (1997). Culturing selects for specific genotypes of *Borrelia burgdorferi* in an enzootic cycle in Colorado. *J Clin Microbiol* **35**, 2359–2364.

Norris, D. E., Johnson, B. J., Piesman, J., Maupin, G. O., Clark, J. L. & Black, W. C., IV (1999). Population genetics and phylogenetic analysis of Colorado *Borrelia burgdorferi*. *Am J Trop Med Hyg* **60**, 699–707.

Ochman, H., Lawrence, J. G. & Groisman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* **405**, 299–304.

Ornstein, K., Berglund, J., Nilsson, I., Norrby, R. & Bergstrom, S. (2001). Characterization of Lyme borreliosis isolates from patients with erythema migrans and neuroborreliosis in southern Sweden. *J Clin Microbiol* **39**, 1294–1298.

Posada, D. & Crandall, K. A. (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**, 817–818.

Posada, D. & Crandall, K. A. (2002). The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **54**, 396–402.

Postic, D., Assous, M. V., Grimont, P. A. & Baranton, G. (1994). Diversity of *Borrelia burgdorferi* sensu lato evidenced by restriction fragment length polymorphism of *rrf* (5S)–*rrl* (23S) intergenic spacer amplicons. *Int J Syst Bacteriol* **44**, 743–752.

Qiu, W. G., Bosler, E. M., Campbell, J. R., Ugine, G. D., Wang, I. N., Luft, B. J. & Dykhuizen, D. E. (1997). A population genetic study of *Borrelia burgdorferi* sensu stricto from eastern Long Island, New York, suggested frequency-dependent selection, gene flow and host adaptation. *Hereditas* **127**, 203–216.

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

Rozas, J. & Rozas, R. (1999). DnaSP version 3: an integrated program for molecular population genetics and molecular evolution analysis. *Bioinformatics* **15**, 174–175.

Sawyer, S. (1989). Statistical tests for detecting gene conversion. *Mol Biol Evol* **6**, 526–538.

Schwan, T. G., Piesman, J., Golde, W. T., Dolan, M. C. & Rosa, P. A. (1995). Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *Proc Natl Acad Sci U S A* **92**, 2909–2913.

Scoles, G. A., Papero, M., Beati, L. & Fish, D. (2001). A relapsing fever group spirochete transmitted by *Ixodes scapularis* ticks. *Vector Borne Zoonotic Dis* **1**, 21–34.

Seinost, G., Dykhuizen, D. E., Dattwyler, R. J., Golde, W. T., Dunn, J. J., Wang, I. N., Wormser, G. P., Schriefer, M. E. & Luft, B. J. (1999a). Four clones of *Borrelia burgdorferi* sensu stricto cause invasive infection in humans. *Infect Immun* **67**, 3518–3524.

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

Steere, A. C. (2001). Lyme disease. *N Engl J Med* **345**, 115–125.

Steere, A. C., Grodzicki, R. L., Kornblatt, A. N., Craft, J. E., Barbour, A. G., Burgdorfer, W., Schmid, G. P., Johnson, E. & Malawista, S. E. (1983). The spirochetal etiology of Lyme disease. *N Engl J Med* **308**, 733–740.

Swofford, D. L. (2001). *PAUP. Phylogenetic Analysis Using Parsimony (* and other methods). Beta*, version 4.0b6. Sunderland, MA: Sinauer Associates.

Theisen, M., Borre, M., Mathiesen, M. J., Mikkelsen, B., Lebech, A. M. & Hansen, K. (1995). Evolution of the *Borrelia burgdorferi* outer surface protein OspC. *J Bacteriol* **177**, 3036–3044.

Thompson, J. D., Higgins, D. G. & Gibson, T. J. (1994). CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* **22**, 4673–4680.

Tsao, J. (2000). Vertebrate host community composition and the dynamics of *Borrelia burgdorferi*, the etiologic agent of Lyme disease: theory and experiments. In *Department of Ecology and Evolution*, p. 90. Chicago: University of Chicago.

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

Wilske, B., Preac-Mursic, V., Schierz, G. & Busch, K. V. (1986). Immunochemical and immunological analysis of European *Borrelia burgdorferi* strains. *Zentbl Bakteriol Mikrobiol Hyg [A]* **263**, 92–102.

Wilske, B., Preac-Mursic, V., Gobel, U. B., Graf, B., Jauris, S., Soutschek, E., Schwab, E. & Zumstein, G. (1993). An OspA serotyping system for *Borrelia burgdorferi* based on reactivity with monoclonal antibodies and OspA sequence analysis. *J Clin Microbiol* **31**, 340–350.

Wilske, B., Jauris-Heipke, S., Lobentanzer, R., Pradel, I., Preac-Mursic, V., Rossler, D., Soutschek, E. & Johnson, R. C. (1995). Phenotypic analysis of outer surface protein C (OspC) of *Borrelia burgdorferi* sensu lato by monoclonal antibodies: relationship to genospecies and OspA serotype. *J Clin Microbiol* **33**, 103–109.

Wormser, G. P., Liveris, D., Nowakowski, J., Nadelman, R. B., Cavaliere, L. F., McKenna, D., Holmgren, D. & Schwartz, I. (1999). Association of specific subtypes of *Borrelia burgdorferi* with hematogenous dissemination in early Lyme disease. *J Infect Dis* **180**, 720–725.