# MLST of housekeeping genes captures geographic population structure and suggests a European origin of *Borrelia burgdorferi*

Gabriele Margos[a,b], Anne G. Gatewood[c], David M. Aanensen[d], Klára Hanincová[e], Darya Terekhova[e], Stephanie A. Vollmer[a], Muriel Cornet[f], Joseph Piesman[g], Michael Donaghy[h], Antra Bormane[i], Merrilee A. Hurn[j], Edward J. Feil[a], Durland Fish[c], Sherwood Casjens[k], Gary P. Wormser[l], Ira Schwartz[e], and Klaus Kurtenbach[a]

Departments of [a]Biology and Biochemistry and [j]Mathematical Sciences, University of Bath, Bath BA2 7AY, United Kingdom; [c]Department of Epidemiology and Public Health, Yale School of Medicine, Yale University, New Haven, CT 06520; [d]Department of Infectious Disease Epidemiology, Imperial College London, St. Mary's Hospital, London W2 1PG, United Kingdom; [e]Department of Microbiology and Immunology and [l]Division of Infectious Diseases, Department of Medicine, New York Medical College, Valhalla, NY 10595; [f]Centres Nationaux de Référence des Borrelia et de la Leptospirose, Institut Pasteur, 75724 Paris Cedex 15, France; [g]Bacterial Diseases Branch, Division of Vector-Borne Infectious Diseases, Centers for Disease Control and Prevention, Ft. Collins, CO 80521; [h]Department of Clinical Neurology, John Radcliffe Hospital, University of Oxford, Oxford OX3 9DU, United Kingdom; [i]Public Health Agency, LV-1012, Riga, Latvia; and [k]Division of Cell Biology and Immunology, Department of Pathology, University of Utah Medical School, Salt Lake City, UT 84132

Edited by Barry J. Beaty, Colorado State University, Fort Collins, CO, and approved April 23, 2008 (received for review January 16, 2008)

**Lyme borreliosis, caused by the tick-borne bacterium *Borrelia burgdorferi*, has become the most common vector-borne disease in North America over the last three decades. To understand the dynamics of the epizootic spread and to predict the evolutionary trajectories of *B. burgdorferi*, accurate information on the population structure and the evolutionary relationships of the pathogen is crucial. We, therefore, developed a multilocus sequence typing (MLST) scheme for *B. burgdorferi* based on eight chromosomal housekeeping genes. We validated the MLST scheme on *B. burgdorferi* specimens from North America and Europe, comprising both cultured isolates and infected ticks. These data were compared with sequences for the commonly used genetic markers *rrs-rrlA* intergenic spacer (IGS) and the gene encoding the outer surface protein C (*ospC*). The study demonstrates that the concatenated sequences of the housekeeping genes of *B. burgdorferi* provide highly resolved phylogenetic signals and that the housekeeping genes evolve differently compared with the IGS locus and *ospC*. Using sequence data, the study reveals that North American and European populations of *B. burgdorferi* correspond to genetically distinct populations. Importantly, the MLST data suggest that *B. burgdorferi* originated in Europe rather than in North America as proposed previously.**

evolution | Lyme borreliosis | ticks

Lyme borreliosis is the most prevalent vector-borne disease in the temperate zone of the Northern Hemisphere. It is a tick-borne bacterial zoonosis, with hard ticks of the genus *Ixodes* acting as vectors (1) and various vertebrate species serving as reservoir hosts (2–6). At present, Lyme borreliosis spirochetes constitute a group of 13 named species (7–18). Several of these, namely *Borrelia burgdorferi*, *Borrelia afzelii*, *Borrelia garinii*, and *Borrelia spielmanii*, are associated with disease in humans (8, 19–22). *B. burgdorferi* (sometimes referred to as *B. burgdorferi* sensu stricto) occurs in both Europe and North America (23). In the United States, Lyme disease is caused by *B. burgdorferi* only (24).

Despite almost three decades of research and control efforts, the numbers of new human cases of Lyme borreliosis in the United States continue to increase (24). It is likely that this reflects increasing population sizes and geographic ranges of pathogenic genotypes of *B. burgdorferi* (25) because it is known that populations of the principal vector tick in eastern North America, *Ixodes scapularis*, have been spreading from past refuges into new woodland habitats (26–34). In addition, it is likely that the wide host range of *B. burgdorferi* has been facilitating its epidemic dispersal in the northeastern United States (2, 3). There are three main regions in the United States where *B. burgdorferi* is now prevalent: the Northeast, the upper Midwest, and northern coastal California. *B. burgdorferi* is relatively rare in Europe compared with the northeastern United States and has not been recorded in Asia (4, 35–37).

The genome of *B. burgdorferi* is remarkable among bacteria, in that it consists of a linear chromosome, which encodes ≈50% of the predicted proteome, and a large number of linear and circular plasmids (38, 39). Both chromosomal and plasmid-located loci have been used for typing *B. burgdorferi*, such as the chromosomal *rrs-rrlA* intergenic spacer (IGS) (40, 41), the gene encoding the membrane protein p66 (42), or the plasmid-located, hypervariable gene encoding the immunodominant outer surface protein C (*ospC*) (25, 43).

Unambiguous genotyping systems are key to describing epidemiological and ecological patterns and illuminating the evolutionary processes that shape microbial populations. Several recent studies have used sequence information of multiple loci to characterize Lyme borreliosis spirochetes (15, 17, 41, 44, 45). However, these typing approaches deviate from typical multilocus sequence typing (MLST) schemes or multilocus sequence analysis (MLSA) (46) developed for other microbial pathogens in that different categories of loci were combined, such as hypervariable genes encoding outer surface proteins, conserved housekeeping genes, or noncoding loci.

Most MLST/MLSA schemes are based on housekeeping genes, which are subject to purifying selection and slow evolution, and the variation within these genes is nearly neutral (47). Although there are normally fewer polymorphic sites in individual housekeeping genes compared with hypervariable genes, using the combined sequences of multiple housekeeping genes

Author contributions: G.M., D.F., and K.K. designed research; G.M., A.G.G., K.H., and S.A.V. performed research; A.G.G., D.M.A., K.H., D.T., M.C., J.P., M.D., A.B., S.C., and G.P.W. contributed new reagents/analytic tools; G.M., M.A.H., E.J.F., I.S., and K.K. analyzed data; and G.M., E.J.F., I.S., and K.K. wrote the paper.

The authors declare no conflict of interest.

This article is a PNAS Direct Submission.

Freely available online through the PNAS open access option.

Data deposition: The sequences reported in this paper have been deposited in the GenBank database (accession nos. EU375814–EU375823 and EU377743–EU377781 for *ospC* and EU375824–EU375833 and EU377782–EU377822 for the IGS). Sequences of the housekeeping genes have been submitted to the Multi Locus Sequence Typing web site, www.mlst.net and can be accessed via strain ID or ST.

[b]To whom correspondence should be addressed. E-mail: gm250@bath.ac.uk.

This article contains supporting information online at www.pnas.org/cgi/content/full/0800323105/DCSupplemental.

has been shown to provide high discriminatory power while retaining signatures of longer-term evolutionary relationships or clonal stability (47–50) (see also www.mlst.net). Furthermore, analyses of multiple loci can buffer against potentially skewed evolutionary pictures obtained by single-locus analyses (49, 51). Despite their power, typical MLST/MLSA schemes have not yet been applied as tools in population or landscape genetics studies of vector-borne pathogens.

To contribute to the understanding of the emergence of Lyme borreliosis, we aimed to test whether *B. burgdorferi* is structured geographically and to infer the evolutionary origin of this bacterial species. We, therefore, have developed a MLST scheme for *B. burgdorferi* based on housekeeping genes to characterize *B. burgdorferi* populations at the different phylogenetic levels required for evolutionary, epidemiological, and population genetics analyses. We show that North American and European populations of *B. burgdorferi* constitute distinct lineages and that the housekeeping genes evolve differently compared with the IGS locus and *ospC*. Importantly, phylogenetic analyses of the concatenated housekeeping genes suggest that *B. burgdorferi* originated in Europe.

## Results

***B. burgdorferi* Samples and Housekeeping Genes Analyzed.** Specimens of *B. burgdorferi* from the Northeast and Midwest United States, California, and Europe that were analyzed by MLST are listed in supporting information (SI) Table S1. These included field-collected infected ticks and isolates from patients and ticks. The characteristics of the eight housekeeping genes selected, i.e., *clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, and *uvrA*, such as GC content (10) or dN/dS ratio, are shown in Table S2.

**MLST of *B. burgdorferi* and Comparison with the Major IGS Genotypes and *ospC* Major Groups.** In total, 33 sequence types (STs) were defined among the 64 samples of *B. burgdorferi* analyzed (Table 1 and Table S3). Most notably, strains from North America did not share any STs with those from Europe (Fig. S1, Table S1, and Table S3). A permutation test using the allelic profiles of *B. burgdorferi* (Table S3) provided significant evidence ($P < 0.003$) that the *B. burgdorferi* populations from Europe and the United States constitute distinct lineages. The distribution of the dissimilarity statistic is shown in Fig. S2. However, STs 20–23 from Europe (IGS genotype 1) shared alleles with North American strains (ST1 and ST2, IGS genotype 1) for several housekeeping genes (*clpX*, *pepX*, *pyrG*, and *rplB*) (Fig. S1 and Table S3). The same European STs shared an *uvrA* allele with ST14 from the United States, which has been typed as IGS genotype 6.

Cultured isolates of *B. burgdorferi* representing the previously described nine major IGS genotypes (41) were included in this study. Some isolates belonging to the same major IGS genotypes (e.g., 8 and 9) were found to have identical alleles for all eight housekeeping genes (Table 1 and Table S3), pointing to some clonal frame within *B. burgdorferi*. Among the European strains, three (NE49, IPT193, and IPT198) represented IGS genotype 10 (41), and two (Z41293 and Z41493) were found to be new IGS genotypes. Here these were termed IGS genotypes 11 and 12 (Table 1 and Table S4).

The samples analyzed in the present study represent 14 previously described *ospC* major groups (6, 43, 52). According to the definition introduced by Wang *et al.* (43), alleles of *ospC* differ by >8% between *ospC* major groups but are <2% different within an *ospC* major group. In this study three novel *ospC* major groups were detected among the European isolates (termed M1, N1, and O1; Table 1), owing to 8–16% dissimilarity to their closest neighbors in the phylogenetic trees. For a few infected ticks from Latvia, it was not possible to amplify the IGS locus or *ospC* (Table 1).

As reported earlier (43), some *ospC* major groups were found for both New World and Old World strains of *B. burgdorferi*. For example, *ospC* major group B was found for ST7 (North American strains B373 and MR623), ST20, and ST23 (European strains IPT2, IPT69, IPT191, and IPT137) (Table 1). Furthermore, strains IPT58 (Europe) and 47703UT (United States) possess almost identical *ospC* alleles with only a 1-nt difference in 484 bp analyzed. Both housekeeping gene sequences and IGS sequences of IPT58 cluster with other European strains (Figs. 1, 2, and 3).

**Table 1. STs, IGS types, and *ospC* major groups of *B. burgdorferi***

| Strain | MLST ST | IGS type* | *ospC* group† |
|---|---|---|---|
| B31, BL206, B515, 16812UT | 1 | 1 | A |
| Ca4, Ca5, Ca6 | 2 | 1 | A |
| 297, B504, 498801UT | 3 | 2 | K |
| B509 | 4 | 2 | H |
| Ca_WTB27 | 5 | 2 | A |
| Ca92-0953 | 6 | 2 | H |
| MR623, B373 | 7 | 3 | B |
| B156 | 8 | 4 | F |
| MR661, MR654, 15506UT | 9 | 4 | N |
| Ca_WTB32 | 10 | 4 | F |
| JD1, BL538, BL515, 114311UT | 11 | 5 | C |
| BL522, B356 | 12 | 6 | M |
| Ca92-1337 | 13 | 6 | M |
| MR616, 15912UT | 14 | 6 | G |
| B500 | 15 | 7 | I |
| B331, B361 | 16 | 7 | I |
| Ca92-1096 | 17 | 7 | I |
| B485, MR607, MR662 | 18 | 8 | U |
| N40, B418, MR640, B348, 15903UT | 19 | 9 | E |
| IPT2, IPT69, IPT191 | 20 | 1 | B |
| IPT23, IPT190 | 20 | 1 | M1 |
| 20604LT | 20 | 1 | ND |
| 22521LT | 20 | ND | ND |
| IPT19 | 21 | 1 | Q |
| 21509LT | 21 | ND | ND |
| IPT135 | 22 | 1 | Q |
| IPT137 | 23 | 1 | B |
| IPT39 | 24 | 5 | S |
| IPT58 | 24 | 5 | L |
| IPT193, IPT198 | 25 | 10 | N1 |
| NE49 | 26 | 10 | Q |
| Z41293 | 27 | 11 | O1 |
| Z41493 | 28 | 12 | N1 |
| 47703UT | 29 | 2 | L |
| 51405UT | 30 | 6 | B |
| 48102UT | 31 | 9 | E |
| 519014UT | 32 | 2 | H |
| 20111LT | 33 | 5 | ND |

ND, not done because of the lack of PCR products.
*According to the method described by Bunikis *et al.* (41).
†According to the method described by Wang *et al.* (43) and Seinost *et al.* (52).

**Phylogenetic Analyses.** For the concatenated housekeeping genes, trees were constructed by using MrBayes software (53) (Fig. 3) and the neighbor joining method in MEGA 3.1 (54) (Fig. S3*A*). The trees were rooted with sequences of *B. garinii* strain PBi from Europe as an outgroup. The two MLST trees obtained by Bayesian inference and the neighbor joining method were highly similar and reasonably well supported at deep branches and at terminal nodes. In contrast to highly recombinogenic bacterial species, such as *Streptococcus pneumoniae* (55), where the phylogenetic relationships cannot be inferred reliably, a robust intraspecific phylogenetic signal was found to be present in *B.*

MICROBIOLOGY

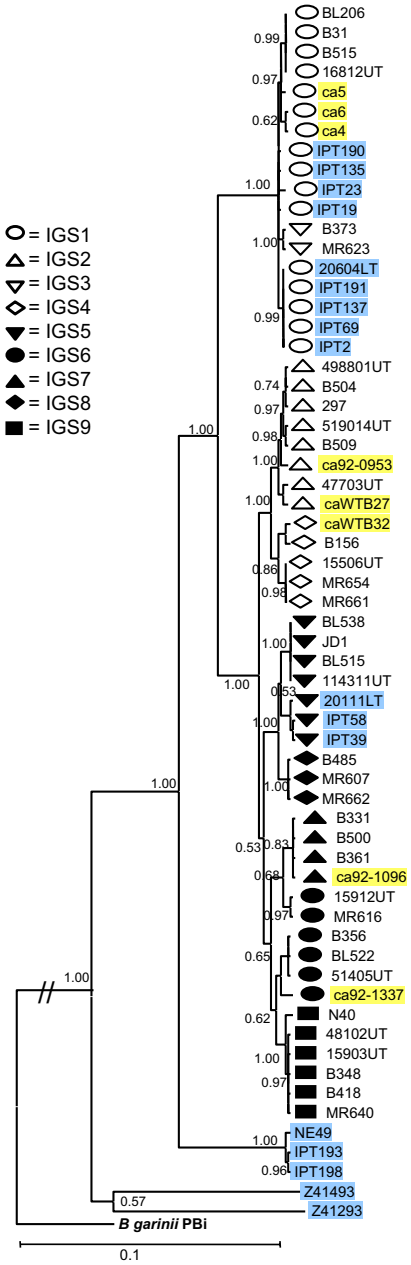


**Fig. 1.** Bayesian phylogenetic inference of IGS sequences of *B. burgdorferi*. Posterior probability values of clades are provided. Symbols refer to the major IGS genotypes as defined by Bunikis *et al*. (41). Samples 21509LT and 22521LT were excluded because of the lack of PCR products. Non-color-coded strains are from the Northeast and Midwest United States. Yellow, strains from California; blue, strains from Europe. The branch length of the outgroup *B. garinii* is not according to scale (indicated by slashes). (Scale bar: 10% divergence.)

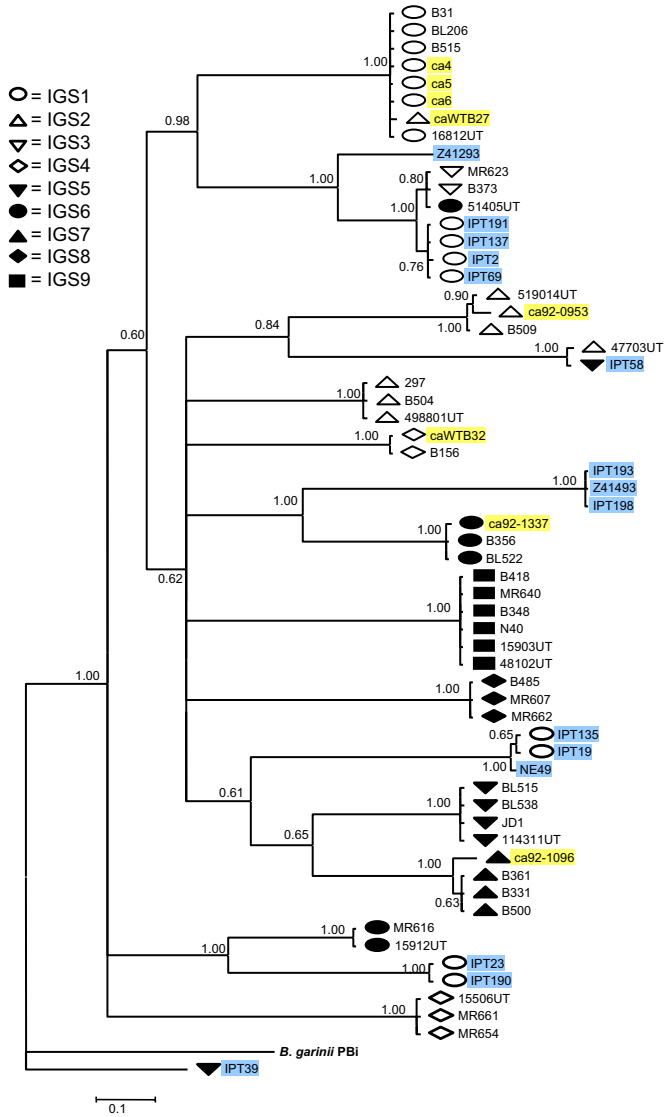


**Fig. 2.** Bayesian phylogenetic inference of *ospC* gene sequences of *B. burgdorferi*. Posterior probability values of clades are provided. Symbols refer to the major IGS genotypes as defined by Bunikis *et al*. (41). Samples 21509LT, 22521LT, 20604LT, and 20111LT are missing because of the lack of PCR products. Non-color-coded strains are from the Northeast and Midwest United States. Yellow, strains from California; blue, strains from Europe. (Scale bar: 10% divergence.)

*burgdorferi*. Except for strains NE49, IPT193, IPT198, Z41293, and Z41493, the European strains formed two subclades among North American strains, but, consistent with the geographical distribution of the STs, North American and European populations of *B. burgdorferi* were placed in distinct terminal nodes. Furthermore, strains from California always branched off from samples derived from the Northeast/Midwest United States (Fig. 3).

The tree based on the IGS locus also revealed patterns of clustering being broadly consistent with geographic source. However, not all spirochetal samples from California and the Northeast/Midwest United States were distinguished in the IGS tree (Fig. 1, i.e., ca92-1096). The topologies of the MLST tree and the IGS tree differ in that several samples representing major IGS genotypes (i.e., IGS genotypes 2, 5, and 6) were divided and placed in different clusters in the MLST tree. A partition homogeneity test (56) using MLST and IGS sequences showed that the sum of the tree length of 99 of 100 random partitions was longer than the sum of the tree length for the original MLST–IGS partition. This confirms that the MLST tree presented here is significantly different from the IGS tree ($P = 0.01$).

The European samples NE49, IPT193, IPT198, Z41293, and Z41493 branched off deeply from the other strains with high posterior probability in both the IGS tree and the MLST tree (Figs. 1 and 3). This divergence was also found for individual housekeeping genes (Fig. S3 *B–I*). The basal position of this

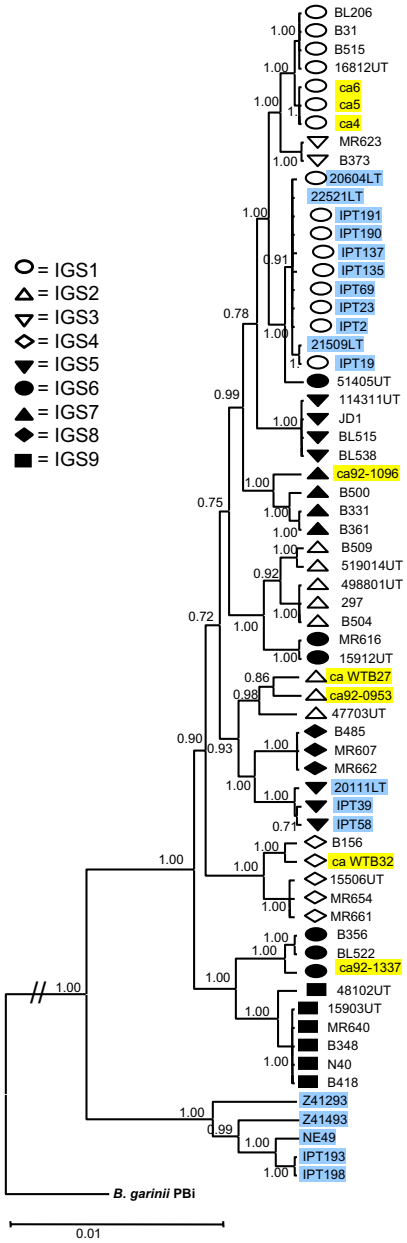


**Fig. 3.** Bayesian phylogenetic inference of concatenated sequences of the housekeeping genes (*clpA*, *clpX*, *nifS*, *pepX*, *pyrG*, *recG*, *rplB*, and *uvrA*) of *B. burgdorferi*. Posterior probability values of clades are provided. Symbols refer to the major IGS genotypes as defined by Bunikis *et al.* (41). Non-color-coded strains are from the Northeast and Midwest United States. Yellow, strains from California; blue, strains from Europe. The branch length of the outgroup *B. garinii* is not according to scale (indicated by slashes). (Scale bar: 1% divergence.)

clade in the MLST tree suggests that these European strains are most closely related to the common ancestor of the strains analyzed in this study.

To test whether the overall differences between the allelic profiles of the European and American strains were due to this diverged European clade, a permutation test was also performed without these strains (i.e., NE49, IPT193, IPT198, Z41293, and Z41493). The results were significant ($P < 0.001$), showing that the overall difference between the European and American populations was not due to this diverged group (data not shown).

The *ospC* tree shows a substantially discordant topology compared with the MLST and IGS trees (Fig. 2). For example, the European samples NE49, IPT193, IPT198, Z41293, and Z41493, which cluster together in the MLST tree, are placed in three different clades scattered across the *ospC* tree. Furthermore, most strains from California were found to cluster together with strains from the Northeast/Midwest United States. Using the *ospC* sequence of *B. garinii* as an outgroup resulted in translocation of IPT39 to the base of the tree (Fig. 2).

## Discussion

Using a sequence-based approach, this study shows that North American and European *B. burgdorferi* populations correspond to distinct lineages. Importantly, the study suggests that *B. burgdorferi* originated in Europe and not in North America as proposed previously based on diversity patterns of *ospC* (57) or arbitrarily primed PCR and pulsed-field gel electrophoresis (58). Comparison of the phylogenetic trees indicates that the commonly used genetic markers of *B. burgdorferi*, the IGS locus and *ospC*, evolve differently compared with the housekeeping genes.

The data presented here using the MLST scheme strongly suggest that it can be used as a powerful means for studying the evolution, epidemiology, and population/landscape genetics of Lyme borreliosis spirochetes. We demonstrate that the approach is capable of (*i*) unambiguously genotyping *B. burgdorferi* samples, (*ii*) establishing evolutionary relationships among the bacterial populations at different levels, (*iii*) capturing geographic population structure of the agents, and (*iv*) typing infections directly from tick vectors without the need for prior culturing of the pathogens.

The clinical samples analyzed in this study were chosen to represent all of the nine previously described major IGS genotypes of *B. burgdorferi* (41). In addition to these clinical samples, the MLST scheme was further validated by using samples of *B. burgdorferi* isolated from ticks and *B. burgdorferi* DNA extracted directly from infected ticks collected in different geographic regions. Whereas the geographic origins of clinical isolates are often not possible to determine, isolates from ticks can be referenced geographically, thereby serving as a useful resource for population studies of *B. burgdorferi*. However, isolation of *B. burgdorferi* may yield a biased picture toward strains that adapt better to *in vitro* conditions (59, 60). Therefore, it is most desirable to genotype *B. burgdorferi* directly from ticks or hosts. In the present study we demonstrate that MLST can be applied equally well to isolates and to ticks infected with *B. burgdorferi* without prior cultivation of the bacteria.

In this study, different STs were determined for North American and European *B. burgdorferi* populations, indicating transcontinental diversification. In contrast, no consistent separation of the regional populations was observed at the level of the major IGS genotypes and *ospC* major groups (Table 1). A signal of geographic resolution similar to that found for the distribution of the STs was captured by the MLST trees, which also discriminated between western and eastern populations of *B. burgdorferi* from North America. However, the sample size of Californian strains used in the present study was small, and more samples need to be analyzed to confirm this observation.

The finding that the regional populations of *B. burgdorferi* differ genetically suggests limited migration of the strains between the regions due to barriers. This is corroborated by the distinct distributional ranges of *I. scapularis*, *Ixodes pacificus*, and *Ixodes ricinus* ticks, the principal vectors of Lyme borreliosis in eastern North America, the Pacific region of North America, and Europe, respectively (5). Although no molecular clock for the evolution of the housekeeping genes of *B. burgdorferi* has yet been established, it is likely that the different populations of *B. burgdorferi* have been separated for a long time (i.e., probably in the order of magnitude of several million years), because housekeeping genes evolve slowly. It is interesting to note that the

MICROBIOLOGY

European samples NE49, IPT193, IPT198, Z41293, and Z41493 formed the most diverged clade among the *B. burgdorferi* samples in the MLST tree. The deep branching of this cluster suggests a common ancestor of the *B. burgdorferi* populations that evolved in Europe. In contrast, based on *ospC*, it has previously been hypothesized that *B. burgdorferi* originated in the New World (57, 61). As discussed below, these conflicting scenarios of speciation and origin of *B. burgdorferi* are likely to be related to different evolutionary pathways of the housekeeping genes and *ospC*.

The shapes of the MLST, IGS, and *ospC* trees as determined in this study differ. First, the topologies of the trees are discordant, with the *ospC* tree being most different. This may indicate that *ospC* has undergone recombination in the past (61). Second, the *ospC* tree is characterized by deeper branching than the MLST and IGS trees, suggesting longer-term evolutionary relationships among the *ospC* major groups or, alternatively, faster evolution. The lack of clear geographic structure of *B. burgdorferi* in North America when using *ospC* as a marker is consistent with earlier studies of *B. burgdorferi* undertaken in the northeastern United States (25) and may be related to the evolution of the *ospC* major groups in ancestral spirochete populations (61). Balancing selection could have maintained such an ancient polymorphism of *ospC* in the wake of past population bottlenecks, as proposed previously (25, 43, 61). In two studies, the European strains NE49, Z41293, and Z41493 have been shown to be borderline as far as the species status is concerned (15, 23). While these strains form distinct clades in the MLST and IGS trees showing significantly increased branch length, they were scattered across the *ospC* tree. The misplacement of these European strains in the *ospC* tree provides indisputable evidence that *ospC* has been evolving differently from the IGS locus and the eight housekeeping genes.

Balancing selection is a form of frequency-dependent selection and should result in a dN/dS ratio of $>1$ for genes under immune selection. However, we and others (43) have found an overall dN/dS ratio of $<1$ for *ospC*, a gene encoding an immunodominant outer surface lipoprotein of *B. burgdorferi*. Sliding window analysis (data not shown) (43) and a "sitewise likelihood-ratio" method (62) showed that different parts of *ospC* display different dN/dS ratios (Table S5). This indicates that some regions of the gene are under positive immune selection whereas others are more conserved because of functional constraints. This is consistent with recent findings that *ospC* encodes a lipoprotein that is essential for colonization of the tick's salivary glands (63, 64) and/or early infection of the vertebrate host (65, 66).

Previous epidemiological and theoretical work suggests that clones and clonal complexes, as demarcated by MLST/MLSA of housekeeping genes, are remarkably congruent with pathotypes and ecotypes in several bacterial species (67). Thus, MLST may also have the power to identify ecotypes within *B. burgdorferi a priori* of determining their ecological niches. This is supported by data from a recent experimental study on the transmissibility of different isolates of *B. burgdorferi* (i.e., BL206 and B348) to ticks by white-footed mice, an important reservoir host in the northeastern United States (68). However, further transmission studies are needed to clarify whether the STs, clones, or clusters identified within *B. burgdorferi* by MLST do consistently correspond to ecotypes that are selectively maintained or, alternatively, whether the lineages defined by MLST evolved through genetic drift (3, 36).

A few other studies have used multiple loci to type Lyme borreliosis spirochetes or to delineate new species within this group as an alternative approach to whole DNA–DNA hybridization (15, 17, 41, 44, 45). These schemes combine different categories of loci that are likely to evolve at different rates. However, using loci that evolve at different rates and under different tree structures may pose problems in inferring phylogenetic trees (69, 70). To provide a genotyping method that can also be used to confidently infer evolutionary pathways, we developed a MLST scheme for *B. burgdorferi* that is based exclusively on housekeeping genes that display an appropriate window of no more than 5% allelic divergence among different strains (71). For relatively clonal bacteria, such as *B. burgdorferi* (72), it is widely accepted that evolutionary relationships can be inferred by using multiple housekeeping genes (73).

Taken together, the study suggests that *B. burgdorferi* originated in Europe but that this species has been prevalent in North America for a long time. It is likely that the surfacing of clinical cases of Lyme borreliosis in North America three decades ago represents the reemergence of this tick-borne pathogen out of refuges in which it has been persisting throughout the post-Columbian settlements and the industrial revolution (3). Future landscape genetic studies of *B. burgdorferi* using MLST as a tool will unveil whether signatures of population bottlenecks, population expansions, and geographic dispersal are present in the housekeeping genes.

## Materials and Methods

*B. burgdorferi* DNA was extracted from cultured isolates and directly from infected ticks. Primers were designed to conserved regions of eight housekeeping genes (Table S6; see also www.mlst.net). These genes as well as the IGS and *ospC* (6, 40, 41) were amplified by nested or seminested PCR followed by DNA sequencing. Sequences of these loci were analyzed for characteristics such as GC content or dN/dS ratio. Sequences of the housekeeping genes were assigned allele numbers according to the MLST web site hosted at Imperial College London (www.mlst.net), and STs were defined. Phylogenetic trees of the individual and concatenated sequences of the housekeeping genes and individual IGS and *ospC* sequences were constructed by using MrBayes software (53) and the neighbor joining method (54). To compare the tree topologies, a partition homogeneity test (56), implemented in PAUP 4.0, was applied. A permutation test was performed to analyze the distributions of the allelic profiles of *B. burgdorferi* from Europe and North America. Details can be found in *SI Materials and Methods*.

GenBank database accession numbers of all IGS and *ospC* sequences used in this study are listed in Table S7.

**ACKNOWLEDGMENTS.** We thank Guy Baranton, Liliana Vitorino, Ronald Jenner, Matthew Wills, Max Telford, Helge Kampen, Elisa Loza Reyes, and Stephen Bent for useful comments and Benjamin Luft for providing unpublished sequence information. This work was funded by the Wellcome Trust (D.M.A. through Brian Spratt, and K.K.) and the National Institutes of Health (K.K., D.F., I.S., and S.C.).

1. Burgdorfer W, *et al.* (1982) Lyme disease—a tick-borne spirochetosis? *Science* 216:1317–1319.
2. Hanincova K, Kurtenbach K, Diuk-Wasser M, Brei B, Fish D (2006) Epidemic spread of Lyme borreliosis, northeastern United States. *Emerg Infect Dis* 12:604–611.
3. Kurtenbach K, *et al.* (2006) Fundamental processes in the evolutionary ecology of Lyme borreliosis. *Nat Rev Microbiol* 4:660–669.
4. Kurtenbach K, Schaefer SM, de Michelis S, Etti S, Sewell HS (2002) in *Lyme Borreliosis: Biology of the Infectious Agents and Epidemiology of Disease*, eds Gray JS, Kahl O, Lane RS, Stanek G (CABI, Wallingford, UK), pp 117–148.
5. Piesman J, Gern L (2004) Lyme borreliosis in Europe and North America. *Parasitology* 129(Suppl):S191–S220.
6. Brisson D, Dykhuizen DE (2004) *ospC* diversity in *Borrelia burgdorferi*: Different hosts are different niches. *Genetics* 168:713–722.
7. Baranton G, *et al.* (1992) Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp nov, and group VS461 associated with Lyme borreliosis. *Int J Syst Bacteriol* 42:378–383.
8. Canica MM, *et al.* (1993) Monoclonal antibodies for identification of *Borrelia afzelii* sp. nov. associated with late cutaneous manifestations of Lyme borreliosis. *Scand J Infect Dis* 25:441–448.
9. Fukunaga M, *et al.* (1996) Characterization of spirochetes isolated from ticks (*Ixodes tanuki*, *Ixodes turdus*, and *Ixodes columnae*) and comparison of the sequences with those of *Borrelia burgdorferi* sensu lato strains. *Appl Environ Microbiol* 62:2338–2344.
10. Johnson RC, Hyde FW, Rumpel CM (1984) Taxonomy of the Lyme disease spirochetes. *Yale J Biol Med* 57:529–537.
11. Kawabata H, Masuzawa T, Yanagihara Y (1993) Genomic analysis of *Borrelia japonica* sp. nov. isolated from *Ixodes ovatus* in Japan. *Microbiol Immunol* 37:843–848.

12. Le Fleche A, Postic D, Girardet K, Peter O, Baranton G (1997) Characterization of *Borrelia lusitaniae* sp. nov. by 16S ribosomal DNA sequence analysis. *Int J Syst Bacteriol* 47:921–925.
13. Marconi RT, Liveris D, Schwartz I (1995) Identification of novel insertion elements, restriction fragment length polymorphism patterns, and discontinuous 23S rRNA in Lyme disease spirochetes: Phylogenetic analyses of rRNA genes and their intergenic spacers in *Borrelia japonica* sp. nov. and genomic group 21038 (*Borrelia andersonii* sp. nov.) isolates. *J Clin Microbiol* 33:2427–2434.
14. Masuzawa T, *et al.* (2001) *Borrelia sinica* sp. nov., a Lyme disease-related *Borrelia* species isolated in China. *Int J Syst Evol Microbiol* 51:1817–1824.
15. Postic D, Garnier M, Baranton G (2007) Multilocus sequence analysis of atypical *Borrelia burgdorferi* sensu lato isolates—description of *Borrelia californiensis* sp. nov., and genomospecies 1 and 2. *Int J Med Microbiol* 297:263–271.
16. Postic D, Ras NM, Lane RS, Hendson M, Baranton G (1998) Expanded diversity among Californian borrelia isolates and description of *Borrelia bissettii* sp. nov. (formerly *Borrelia* group DN127). *J Clin Microbiol* 36:3497–3504.
17. Richter D, *et al.* (2006) Delineation of *Borrelia burgdorferi* sensu lato species by multilocus sequence analysis and confirmation of the delineation of *Borrelia spielmanii* sp. nov. *Int J Syst Evol Microbiol* 56:873–881.
18. Wang G, *et al.* (1997) Genetic and phenotypic analysis of *Borrelia valaisiana* sp. nov. (*Borrelia* genomic groups VS116 and M19). *Int J Syst Bacteriol* 47:926–932.
19. Richter D, Schlee DB, Allgower R, Matuschka FR (2004) Relationships of a novel Lyme disease spirochete, *Borrelia spielmani* sp nov, with its hosts in Central Europe. *Appl Environ Microbiol* 70:6414–6419.
20. Steere AC, *et al.* (1983) The spirochetal etiology of Lyme disease. *N Engl J Med* 308:733–740.
21. Wang G, van Dam AP, Dankert J (1999) Phenotypic and genetic characterization of a novel *Borrelia burgdorferi* sensu lato isolate from a patient with Lyme borreliosis. *J Clin Microbiol* 37:3025–3028.
22. Wilske B, *et al.* (1996) Diversity of OspA and OspC among cerebrospinal fluid isolates of *Borrelia burgdorferi* sensu lato from patients with neuroborreliosis in Germany. *Med Microbiol Immunol* 184:195–201.
23. Postic D, *et al.* (1999) Common ancestry of *Borrelia burgdorferi* sensu lato strains from North America and Europe. *J Clin Microbiol* 37:3010–3012.
24. Piesman J (2006) Strategies for reducing the risk of Lyme borreliosis in North America. *Int J Med Microbiol* 296(Suppl 40):17–22.
25. Qiu WG, Dykhuizen DE, Acosta MS, Luft BJ (2002) Geographic uniformity of the Lyme disease spirochete (*Borrelia burgdorferi*) and its shared history with tick vector (*Ixodes scapularis*) in the Northeastern United States. *Genetics* 160:833–849.
26. Brownstein JS, Holford TR, Fish D (2003) A climate-based model predicts the spatial distribution of the Lyme disease vector *Ixodes scapularis* in the United States. *Environ Health Perspect* 111:1152–1157.
27. Dennis DT, Nekomoto TS, Victor JC, Paul WS, Piesman J (1998) Reported distribution of *Ixodes scapularis* and *Ixodes pacificus* (Acari: Ixodidae) in the United States. *J Med Entomol* 35:629–638.
28. Falco RC, Daniels TJ, Fish D (1995) Increase in abundance of immature *Ixodes scapularis* (Acari: Ixodidae) in an emergent Lyme disease endemic area. *J Med Entomol* 32:522–526.
29. Fish D (1993) in *Ecology and Environmental Management of Lyme Disease*, ed Ginsberg HS (Rutgers Univ Press, New Brunswick, NJ), pp 25–42.
30. Glavanakov S, *et al.* (2001) Lyme disease in New York State: Spatial pattern at a regional scale. *Am J Trop Med Hyg* 65:538–545.
31. Kitron U, Kazmierczak JJ (1997) Spatial analysis of the distribution of Lyme disease in Wisconsin. *Am J Epidemiol* 145:558–566.
32. Madhav NK, Brownstein JS, Tsao JI, Fish D (2004) A dispersal model for the range expansion of blacklegged tick (Acari: Ixodidae). *J Med Entomol* 41:842–852.
33. Spielman A (1994) The emergence of Lyme disease and human babesiosis in a changing environment. *Ann NY Acad Sci* 740:146–156.
34. White DJ, *et al.* (1991) The geographic spread and temporal increase of the Lyme disease epidemic. *J Am Med Assoc* 266:1230–1236.
35. Hubalek Z, Halouzka J (1997) Distribution of *Borrelia burgdorferi* sensu lato genomic groups in Europe, a review. *Eur J Epidemiol* 13:951–957.
36. Kurtenbach K, *et al.* (2002) Host association of *Borrelia burgdorferi* sensu lato—the key role of host complement. *Trends Microbiol* 10:74–79.
37. Kurtenbach K, *et al.* (2001) Distinct combinations of *Borrelia burgdorferi* sensu lato genospecies found in individual questing ticks from Europe. *Appl Environ Microbiol* 67:4926–4929.
38. Casjens S (2000) Borrelia genomes in the year 2000. *J Mol Microbiol Biotechnol* 2:401–410.
39. Casjens S, *et al.* (2000) A bacterial genome in flux: The twelve linear and nine circular extrachromosomal DNAs in an infectious isolate of the Lyme disease spirochete *Borrelia burgdorferi*. *Mol Microbiol* 35:490–516.
40. 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.
41. Bunikis J, *et al.* (2004) Sequence typing reveals extensive strain diversity of the Lyme borreliosis agents *Borrelia burgdorferi* in North America and *Borrelia afzelii* in Europe. *Microbiology* 150:1741–1755.
42. Bunikis J, Noppa L, Ostberg Y, Barbour AG, Bergstrom S (1996) Surface exposure and species specificity of an immunoreactive domain of a 66-kilodalton outer membrane protein (P66) of the *Borrelia* spp. that cause Lyme disease. *Infect Immun* 64:5111–5116.
43. Wang IN, *et al.* (1999) Genetic diversity of *ospC* in a local population of *Borrelia burgdorferi* sensu stricto. *Genetics* 151:15–30.
44. Qiu WG, *et al.* (2004) Genetic exchange and plasmid transfers in *Borrelia burgdorferi* sensu stricto revealed by three-way genome comparisons and multilocus sequence typing. *Proc Natl Acad Sci USA* 101:14150–14155.
45. Attie O, *et al.* (2007) Co-evolution of the outer surface protein C gene (*ospC*) and intraspecific lineages of *Borrelia burgdorferi* sensu stricto in the northeastern United States. *Infect Genet Evol* 7:1–12.
46. Gevers D, *et al.* (2005) Opinion: Re-evaluating prokaryotic species. *Nat Rev Microbiol* 3:733–739.
47. Enright MC, Spratt BG (1999) Multilocus sequence typing. *Trends Microbiol* 7:482–487.
48. Feil EJ, Enright MC (2004) Analyses of clonality and the evolution of bacterial pathogens. *Curr Opin Microbiol* 7:308–313.
49. Maiden MC, *et al.* (1998) Multilocus sequence typing: A portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* 95:3140–3145.
50. Spratt BG (1999) Multilocus sequence typing: Molecular typing of bacterial pathogens in an era of rapid DNA sequencing and the Internet. *Curr Opin Microbiol* 2:312–316.
51. Urwin R, Maiden MC (2003) Multi-locus sequence typing: A tool for global epidemiology. *Trends Microbiol* 11:479–487.
52. Seinost G, *et al.* (1999) Four clones of *Borrelia burgdorferi* sensu stricto cause invasive infection in humans. *Infect Immun* 67:3518–3524.
53. Huelsenbeck JP, Ronquist F (2005) in *Statistical Methods in Molecular Evolution*, ed Nielsen R (Springer, New York), pp 183–232.
54. Kumar S, Tamura K, Nei M (2004) Integrated software for Molecular Evolutionary Genetics Analysis and sequence alignment. *Brief Bioinform* 5:150–163.
55. Feil EJ, Smith JM, Enright MC, Spratt BG (2000) Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 154:1439–1450.
56. Farris JS, Källersjö, M, Kluge AG, Bult C (1995) Constructing a significance test for incongruence. *Syst Biol* 44:570–572.
57. Marti Ras N, Postic D, Foretz M, Baranton G (1997) Borrelia burgdorferi sensu stricto, a bacterial species ''made in the U.S.A.''? *Int J Syst Bacteriol* 47:1112–1117.
58. Foretz M, Postic D, Baranton G (1997) Phylogenetic analysis of *Borrelia burgdorferi* sensu stricto by arbitrarily primed PCR and pulsed-field gel electrophoresis. *Int J Syst Bacteriol* 47:11–18.
59. Norris DE, *et al.* (1997) Culturing selects for specific genotypes of *Borrelia burgdorferi* in an enzootic cycle in Colorado. *J Clin Microbiol* 35:2359–2364.
60. Liveris D, *et al.* (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.
61. Dykhuizen DE, Baranton G (2001) The implications of a low rate of horizontal transfer in *Borrelia*. *Trends Microbiol* 9:344–350.
62. Massingham T, Goldman N (2005) Detecting amino acid sites under positive selection and purifying selection. *Genetics* 169:1753–1762.
63. Fingerle V, Goettner G, Gern L, Wilske B, Schulte-Spechtel U (2007) Complementation of a *Borrelia afzelii* OspC mutant highlights the crucial role of OspC for dissemination of *Borrelia afzelii* in *Ixodes ricinus*. *Int J Med Microbiol* 297:97–107.
64. Pal U, *et al.* (2004) OspC facilitates *Borrelia burgdorferi* invasion of *Ixodes scapularis* salivary glands. *J Clin Invest* 113:220–230.
65. Stewart PE, *et al.* (2006) Delineating the requirement for the *Borrelia burgdorferi* virulence factor OspC in the mammalian host. *Infect Immun* 74:3547–3553.
66. Tilly K, Bestor A, Jewett MW, Rosa P (2007) Rapid clearance of Lyme disease spirochetes lacking OspC from skin. *Infect Immun* 75:1517–1519.
67. Cohan FM (2002) What are bacterial species? *Annu Rev Microbiol* 56:457–487.
68. Hanincova K, *et al.* (2007) Fitness variation of *Borrelia burgdorferi* sensu stricto strains in mice. *Appl Environ Microbiol* 74:153–157.
69. Cunningham CW (1997) Can tree incongruence tests predict when data should be combined? *Mol Biol Evol* 14:733–740.
70. Matsen, FA, Mossel, E, Steel, M (2007) Mixed-up trees: The structure of phylogenetic mixtures. *Bull Math Biol* 70:1115–1139.
71. Terekhova D, Iyer R, Wormser GP, Schwartz I (2006) Comparative genome hybridization reveals substantial variation among clinical isolates of *Borrelia burgdorferi* sensu stricto with different pathogenic properties. *J Bacteriol* 188:6124–6134.
72. Dykhuizen DE, *et al. (1993) Borrelia burgdorferi* is clonal: Implications for taxonomy and vaccine development. *Proc Natl Acad Sci USA* 90:10163–10167.
73. Spratt BG, Hanage WP, Li B, Aanensen DM, Feil EJ (2004) Displaying the relatedness among isolates of bacterial species—the eBURST approach. *FEMS Microbiol Lett* 241:129–134.