

Supporting Materials and Methods

Genome Sequencing. The JD1 and N40 genomes were sequenced by using a shotgun strategy, as described (1, 2). We obtained $\approx 20,000$ sequences to provide approximately eight times coverage of each of the ≈ 1.5 -mb *Borrelia burgdorferi* genomes, following the algorithm of Lander and Waterman (3). Contigs were assembled into genome scaffolds using a series of gap closure strategies, including the construction and sequencing the ends of BAC and other large-insert clones and designing outward primers for the small- or medium-insert random libraries. We identified the putative genes using the software GLIMMER (4). Putative functions of ORFs are according to the original annotation for *B. burgdorferi* B31 (www.tigr.org/tigr-scripts/CMR2/CMRHomePage.spl). We used TMHMM (5) to identify potential membrane-spanning domains, and SIGNALP (6) to identify potential cleavage sites in secreted proteins (based all on B31 sequences). Also, we used the Casjens *et al.* (2) supplementary material list of putative lipoproteins.

Identification of Orthologous ORF Pairs. Scaffold sequences from JD1 and N40 were aligned with finished B31 genome sequences by using the NUCMER program of the MUMMER package (4). Two scaffolds in each of the JD1 and N40 genome assemblies uniquely match the B31 main chromosome without ambiguity. Due to the highly paralogous nature of B31 plasmids (2) some JD1 and N40 contigs often align with multiple plasmid sequences. Of the 21 B31 plasmids, we were able to identify N40 scaffolds that uniquely match the cp26, lp25, cp9, lp17, lp28-2, lp28-4, lp36, lp38, and lp54 plasmids of the B31 genome. Scaffolds that match other plasmids are either missing (lp21, lp5, lp28-1) or nonunique (cp32-1, -3, -4, -6, -7, -8, -9, lp56). We regard the unique NUCMER matches between JD1/N40 scaffolds and the B31 genome as the orthologous contigs to each other. For each pair of orthologous genome segments, we identified nonoverlapping long (≥ 250 nt) ORFs on both genomes by using GLIMMER (4). ORFs in syntenic relation in two orthologous genome segments are considered ortholog ORFs (e.g., Fig. 4).

Multilocus Sequence Typing. *Clinical isolates.* We performed multilocus genomic typing on a collection of *B. burgdorferi sensu stricto* clinical isolates from the northeastern United States by using a library of most variable polymorphic loci identified through comparative genomics. *B. burgdorferi* strains used in this study are described in Table 3. Strains from the skin of primary erythema migrans lesions or blood of patients with Lyme disease were isolated as described (7–9) or were obtained from the Centers for Disease Control. For *in vitro* propagation, spirochetes were cultivated at 34°C in complete BSK-H medium (Sigma).

DNA isolation. For isolation of genomic DNA, 10 ml of low-passage log-phase bacteria were harvested by centrifugation at 10,000 rpm for 30 min at 4°C. The bacterial pellet was washed twice with Tris•Cl buffer (10 mM Tris, pH 7.5/100 mM NaCl), and resuspended in 430 µl of TES (10 mM Tris, pH 7.5/100 mM NaCl/10 mM EDTA). We then added 10 µl of freshly prepared lysozyme (50 mg/ml), 50 µl of Sarkosyl (10%), and 10 µl of proteinase K (10 mg/ml). The mixture was incubated at 50°C overnight before RNase treatment. DNA was extracted with phenol/chloroform once and chloroform once, precipitated with ethanol, and resuspended in TE buffer (1 mM Tris, pH 7.5/1 mM EDTA).

Amplification and sequencing of selected loci. The *ospC* genes from all isolates were amplified by PCR and sequenced as previously described (10, 11). Major *ospC* groups were determined by using the computer programs BLASTN and BLASTX through GenBank at the National Center for Biotechnology Information (NCBI). Primers for other loci were designed by using conserved regions based on comparisons of three genomes after performing localized CLUSTALW alignments. All primers used for this study are presented in Table 4. PCR amplification was performed in 50 µl containing 200 mM each deoxynucleoside triphosphate, 2.0 mM MgSO₄, 2.5 units of Platinum *Taq*DNA polymerase High Fidelity (Invitrogen), 0.5 µM each primer and 100 ng of genomic DNA template. After a denaturation step at 94°C for 1 min, samples underwent 30 cycles of denaturation at 94°C for 30s, annealing at 55°C for 30s, and extension at 68°C for 1.5 min and a final extension step at 68°C for 10 min. PCR products were purified by GFX

chromatography (Amersham Pharmacia Biotech), resolved by agarose gel electrophoresis and visualized by ethidium bromide staining. Purified amplicons were sequenced by using standard dideoxy terminator chemistry as outlined below by using the forward and reverse PCR primers. Absence of specific PCR products, indicating potential absence of particular genetic loci, was confirmed by follow-up amplifications of the flanking DNA segments.

DNA sequencing. Automated DNA sequencing was performed by the State University of New York at Stony Brook Core DNA Sequencing Facility of both strands of each fragment by the dye-terminator method by using the same oligonucleotide primers used for PCR amplification or, where required, appropriate internal primers. Sequences were inspected and assembled with the aid of the SEQUENCHER program (Life Codes). DNA sequences were analyzed by using the BLASTN program through GenBank at NCBI. Nucleotide and protein sequence alignments were performed with MACVECTOR (version 6.5, Oxford Molecular Group).

1. Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., White, O., Ketchum, K. A., Dodson, R., Hickey, E. K., *et al.* (1997) *Nature* **390**, 580–586.
2. Casjens, S., Palmer, N., van Vugt, R., Huang, W. M., Stevenson, B., Rosa, P., Lathigra, R., Sutton, G., Peterson, J., Dodson, R. J., *et al.* (2000) *Mol. Microbiol.* **35**, 490–516.
3. Lander, E. S. & Waterman, M. S. (1988) *Genomics* **2**, 231–239.
4. Delcher, A. L., Phillippy, A., Carlton, J. & Salzberg, S. L. (2002) *Nucleic Acids Res.* **30**, 2478–2483.
5. Sonnhammer, E. L., von Heijne, G. & Krogh, A. (1998) *Proc. Int. Conf. Intell. Syst. Mol. Biol.* **6**, 175–182.
6. Nielsen, H., Engelbrecht, J., Brunak, S. & von Heijne, G. (1997) *Protein Eng* **10**, 1–6.
7. Barbour, A. G. (1984) *Yale J Biol Med* **57**, 521–525.
8. Berger, B. W., Johnson, R. C., Kodner, C. & Coleman, L. (1992) *J. Clin. Microbiol.* **30**, 359–361.

9. Wormser, G. P., Nowakowski, J., Nadelman, R. B., Bittker, S., Cooper, D. & Pavia, C. (1998) *J. Clin. Microbiol.* **36**, 296–298.
10. Wang, I. N., Dykhuizen, D. E., Qiu, W., Dunn, J. J., Bosler, E. M. & Luft, B. J. (1999) *Genetics* **151**, 15–30.
11. 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. (1999) *Infect Immun* **67**, 3518–3524.