

A Genome-Wide Proteome Array Reveals a Limited Set of Immunogens in Natural Infections of Humans and White-Footed Mice with *Borrelia burgdorferi*^{∇†}

Alan G. Barbour,^{1,2,3*} Algimantas Jasinskas,^{1,2,3,‡} Matthew A. Kayala,^{4,5,‡} D. Huw Davies,^{1,3,6}
Allen C. Steere,⁷ Pierre Baldi,^{4,5} and Philip L. Felgner^{1,3,6}

Pacific Southwest Center for Research on Emerging Infections,¹ Department of Microbiology and Molecular Genetics,² Department of Medicine,³ Department of Computer Science,⁴ and Institute of Genomics and Bioinformatics,⁵ University of California Irvine, Irvine, California; ImmPORT Therapeutics Inc., Irvine, California⁶; and Center for Immunology and Inflammatory Diseases, Massachusetts General Hospital, Harvard Medical School, Boston, Massachusetts⁷

Received 14 January 2008/Returned for modification 25 February 2008/Accepted 5 May 2008

Humans and other animals with Lyme borreliosis produce antibodies to a number of components of the agent *Borrelia burgdorferi*, but a full accounting of the immunogens during natural infections has not been achieved. Employing a protein array produced in vitro from 1,292 DNA fragments representing ~80% of the genome, we compared the antibody reactivities of sera from patients with early or later Lyme borreliosis to the antibody reactivities of sera from controls. Overall, ~15% of the open reading frame (ORF) products (Orfs) of *B. burgdorferi* in the array detectably elicited an antibody response in humans with natural infections. Among the immunogens, 103 stood out on the basis of statistical criteria. The majority of these Orfs were also immunogenic with sera obtained from naturally infected *Peromyscus leucopus* mice, a major reservoir. The high-ranking set included several *B. burgdorferi* proteins hitherto unrecognized as immunogens, as well as several proteins that have been established as antigens. The high-ranking immunogens were more likely than nonreactive Orfs to have the following characteristics: (i) plasmid-encoded rather than chromosome-encoded proteins, (ii) a predicted lipoprotein, and (iii) a member of a paralogous family of proteins, notably the Bdr and Erp proteins. The newly discovered antigens included Orfs encoded by several ORFs of the lp36 linear plasmid, such as BBK07 and BBK19, and proteins of the flagellar apparatus, such as FlIL. These results indicate that the majority of deduced proteins of *B. burgdorferi* do not elicit antibody responses during infection and that the limited sets of immunogens are similar for two different host species.

Infectious disease research has advanced rapidly with the accumulation of whole-genome sequences of pathogens and the subsequent use of genome-wide DNA microarrays to study gene expression. Equipped with arrays in different formats, investigators have identified different genes in a variety of pathogens that are more highly expressed in host animals or under in vitro conditions mimicking the in vivo environment. With few exceptions (27), these array studies have been performed with experimental animals, usually rodents, and in laboratory settings. Less is known about the proteins that are expressed during natural infections of humans or other host animals. Detection of a specific antibody during the course of infection is indirect evidence of in vivo expression by the pathogen. But the use of this approach to study large numbers of proteins has been largely limited to one-dimensional and two-dimensional gel electrophoresis of whole cells having an in vitro origin, followed by identification of the more abundant antigens by partial amino acid sequencing of reactive bands or spots and then searches of the databases (22, 23, 38, 44, 52, 60, 66).

An alternative to using the pathogen itself as the source of the proteins is to produce recombinant polypeptides based on the deduced open reading frames (ORFs) of the pathogen's genome and to determine whether these polypeptides are antibody targets (11, 61). A potential shortcoming of using this approach with cells, such as *Escherichia coli* or yeast (*Saccharomyces cerevisiae*) cells, is that some foreign proteins may not be expressed or the quantity may be insufficient. Felgner and coinvestigators increased the success rate for expression and decreased the cost with a high-throughput, cell-free, coupled transcription-translation system (26, 29, 30). Hundreds to thousands of individual recombinant proteins are printed on chips, which are then used to capture antibodies present in serum from infected individuals and other animals. The amount of captured antibody is quantified using a labeled secondary antibody. Whole-proteome microarrays have been employed in studies of experimental *Francisella tularensis* infections in laboratory mice (35, 57) and of immune responses of humans to immunization with live vaccinia virus (26, 29, 31). McKevitt et al. (61) and Brinkmann et al. (11) used the enzyme-linked immunosorbent assay (ELISA) format to study the binding of antibodies of experimentally infected rabbits and people with syphilis to a nearly complete representation of the ORF products (Orfs) of *Treponema pallidum*. Arrays of proteins produced in vitro were also used to characterize antibody responses of humans with tularemia (87) or with *Plasmodium falciparum* malaria (86), but limited sets of selected Orfs were used.

* Corresponding author. Mailing address: University of California Irvine, 3012 Hewitt, Irvine, CA 92697-4028. Phone: (949) 824-5626. Fax: (949) 824-8598. E-mail: abarbour@uci.edu.

† Supplemental material for this article may be found at <http://iai.asm.org/>.

‡ A.J. and M.A.K. contributed equally to this study.

∇ Published ahead of print on 12 May 2008.

Here, we used a genome-wide proteome to characterize antibodies of humans and wild white-footed mice (*Peromyscus leucopus*) naturally infected with the Lyme borreliosis (LB) agent, *Borrelia burgdorferi*. LB is the most frequent tick-borne disease in the northern hemisphere, and its incidence continues to increase in the United States (for a review, see reference 83). A major reservoir for *B. burgdorferi* in the United States is the white-footed mouse; in some areas nearly all mice become infected during the spring and summer (18, 90). In humans, the first manifestations of *B. burgdorferi* infection, typically a solitary rash called erythema migrans, may be followed by manifestations of disseminated infection. These manifestations most commonly involve the joints, heart, or nervous system, as well as skin locations that are distant from the original rash. Dissemination to organs or tissues often requires more intense or longer antibiotic treatment and may be associated with a protracted convalescence in some patients. In a small percentage of patients, pauciarticular arthritis ("Lyme arthritis"), a late manifestation of *B. burgdorferi* infection, persists for months or even several years after antibiotic therapy (82). A commercial vaccine against Lyme disease was available for a few years but was withdrawn from the market (64).

A clinical diagnosis of LB built upon observation of the characteristic skin rash and elicitation of consistent epidemiologic features (e.g., exposure to ticks in an area where the organism is endemic during the season of transmission) has high predictive value (83). But in the absence of a telltale skin rash early in infection or when disseminated forms are suspected, confirmation of suspected LB has largely depended on whole-cell-based serologic tests using an ELISA or Western blot format rather than direct detection or isolation of an etiologic agent (1, 14). According to the most commonly used criterion (33), Western blot positivity for immunoglobulin G (IgG) antibodies calls for at least 5 bands out of a list of 10 bands. Some of the antigens with informative value in *B. burgdorferi* lysates have been identified with certain gene products, but most of these antigens remain known only from their migrations in gels.

Our intent was to use a recombinant proteome representing most of the genome of *B. burgdorferi* to obtain an accounting of the targets of antibodies that is more comprehensive than the accounting which has been feasible up to now for this pathogen in natural infections. We applied an array to analysis of sera from humans with LB and white-footed mice with the aims of discovering new antigens and, by indirect means, identifying proteins that are expressed *in vivo*. To these ends, we designed experiments to obtain information about all the proteins that were immunogenic, i.e., all the proteins that detectably elicited antibodies. We believed that the results would also provide insights about proteins that were not demonstrably immunogenic. One possible outcome of the experiment was a failure to identify any new immunogens of *B. burgdorferi* beyond those that have been identified piecemeal so far. But an equally plausible experimental outcome was finding that a majority of the Orfs of *B. burgdorferi* are immunogenic. The results were between these two extremes; several new antigens were identified, but these antigens belonged to a limited set of proteins that elicited antibody responses during infection. The large majority of Orfs were not detectably immunogenic in humans

with different stages of LB and in rodent reservoirs of *B. burgdorferi* in an area where the organism is endemic.

MATERIALS AND METHODS

Bacterial strain, genome sequences, and primer design. Strain B31 of *B. burgdorferi* had undergone three passages since its isolation by one of us (6, 19). This organism was cultivated in BSK II broth medium (6). A high-passage isolate of strain B31 had been cloned by limiting dilution and had been serially passed in culture medium at least 50 times. Whole-genome DNA was extracted from the low-passage isolate as described previously (62). Primers were based on the sequences and annotations of the chromosome and 21 plasmids of strain B31 (accession numbers NC_001318, NC_000948 to NC_000957, NC_001903, NC_001904, and NC_001849 to NC_001857) (<http://www.blackwellpublishing.com/products/journals/suppmat/mole/casjens.htm>) (21, 39). Forward and reverse primers were 20 nucleotides long and were complementary to the 5' and 3' ends of each ORF; peripherally they also included 33 nucleotide adapter sequences specific for plasmid pXT7 for recombination cloning, as described previously (29). The forward and reverse primers are described at the http://contact14.ics.uci.edu/virus/borrelia_index.php website. We also included the type K OspC protein gene of strain 297, in addition to the type A OspC gene of B31 (12, 15). To name ORFs, we used the designations assigned to strain B31's genome (21, 39); "BB" followed by a four digit number (e.g., BB0279) indicates a chromosome ORF, while "BB" followed by a third letter and a two-digit number (e.g., BBA25) indicates a linear or circular plasmid ORF, and each replicon is assigned a separate letter (e.g., "A" for linear plasmid lp54 or "B" for circular plasmid cp26). As needed, we supplemented genome ORF designations with names in common use or when polypeptide identity has been inferred from homology to proteins with known functions. The predictions of lipoproteins are those of Casjens et al. (<http://www.blackwellpublishing.com/products/journals/suppmat/mole/casjens.htm>).

Array production. PCR amplification, cloning of amplicons into the plasmid vector, and then transformation of *E. coli* DH5 α were carried out as described previously (29, 86). Of the 1,640 ORFs that were identified in the *B. burgdorferi* strain B31 genome (21, 39), 1,513 (861 chromosomal genes and 652 plasmid genes) were subjected to PCR with the specific primers. The remaining 127 ORFs had sequences that were so similar to the sequence of at least one other ORF that PCR primers would not distinguish between them. Of the 861 chromosomal ORFs that we attempted to amplify, 783 (91%) produced a product that was the correct size when PCR was performed, and 756 (88%) were successfully cloned into the vector. Of the 652 plasmid ORFs, 572 (88%) were amplified, and 536 (82%) were cloned into the plasmid vector. A sample consisting of 7% of 1,292 clones from strain B31 was randomly selected for sequencing, and the insert was confirmed in all cases. The coefficient of determination (R^2) between the sizes of the ORFs and cloning success was only 0.05. The following 26 plasmid ORFs were randomly selected to be replicated on the array: BBA03, BBA04, BBA14, BBA25, BBA52, BBA59, BBA62, BBA69, BBB07, BBB19, BBC06, BBJ50, BBK50, BBL28, BBL39, BBM38, BBN37, BBO40, BBP28, BBQ35, BBQ60, BBR80, BBR28, BBR42, BBS30, and BBT07. As a negative control, the arrays also contained 14 pairs of spots with the *E. coli* coupled transcription-translation reaction mixture without plasmid DNA.

Plasmid DNA was extracted and isolated using QIAprep spin kits (Qiagen). *In vitro* coupled transcription-translation reactions were performed with RTS 100 *E. coli* HY kits (Roche) in 0.2-ml tubes that were incubated for 5 h at 30°C. The presence of the polyhistidine tag at the N terminus of the recombinant protein and the presence of the influenza A hemagglutinin at the protein's C terminus were detected with monoclonal antibodies His-1 (Sigma) and 3F10 (Roche), respectively, and confirmed expression in the *in vitro* reactions. Products of transcription-translation reactions were printed in duplicate on nitrocellulose-coated glass slides (FAST; Whatman) using an Omni Grid 100 apparatus (Genomic Solutions).

Protein purification. Plasmid DNA was extracted from selected clones and transformed in *E. coli* BL21 Star(DE3)/pLysS cells as described by the manufacturer (Invitrogen). The resultant transformants were cultivated in Terrific broth (Bio 101 Systems) to stationary phase and, after harvesting by centrifugation, were lysed with BugBuster buffer (Novagen). The lysate was applied to a 5-ml HiTrap chelating HP affinity column (GE Healthcare). After the column was washed, bound proteins were eluted with an imidazole step gradient using an Amersham Biosciences AKTA fast protein liquid chromatography system operated with UNICORN 5.01 software. The average amount recovered from a 1.0-liter culture was 1 to 3 mg of protein with a purity of 80 to 90%, as estimated by sodium dodecyl sulfate-polyacrylamide gel electrophoresis. Recovered proteins were printed on array slides, as described above, or subjected to polyacryl-

amide gel electrophoresis with a 4 to 15% acrylamide gradient and then transferred to nitrocellulose membranes for Western blot analysis (20). For printing on the array, the protein concentrations were 0.03, 0.1, 0.3, and 0.9 mg/ml.

Serum samples. All serum samples were originally collected for other studies for which informed consent had been obtained; patient identifier information had been removed. Serum panel 1 included samples from 48 adults collected between 1990 and 1994, including 24 patients with erythema migrans (early infection), 19 patients with dissemination to other organs or other evidence of persistent infection (later infection), and 5 healthy controls from a region where the organism is not endemic. These samples were provided by the Centers for Disease Control and Prevention, Fort Collins, CO, which had performed flagellin-based ELISA and IgG and IgM Western blot assays, as described previously (17, 50). Panel 1 also included sera from 13 healthy adult volunteers residing in California.

Serum panel 2 included serum specimens from 20 healthy adult control subjects, 20 adult patients with culture-positive erythema migrans (early infection), and 20 individuals with persistent LB with oligoarticular arthritis (later infection). All 40 patients with Lyme disease met the criteria of the Centers for Disease Control and Prevention for diagnosis of Lyme disease (95). The 20 patients with erythema migrans were a random sample of 93 patients, seen in a study of early LB, from whom *B. burgdorferi* was cultured from erythema migrans skin lesions (84, 93). Only convalescent samples, which were obtained at the conclusion of 3 to 4 weeks of antibiotic therapy, were tested for these patients, because seropositivity is more frequent during convalescence than during acute infection. The 20 patients with Lyme arthritis were seen in 2006 and early 2007 in a study of susceptibility to Lyme arthritis (82). For 10 of the 20 patients with Lyme arthritis there was resolution of arthritis with antibiotic therapy (antibiotic-responsive arthritis), and for 10 there was not resolution (antibiotic-refractory arthritis). The samples were obtained when the patients had active arthritis. All serum samples had been kept frozen at -80°C until use.

Sera from 10 white-footed mice (*P. leucopus*), which had been captured and then released after blood samples had been obtained at a field site in Connecticut under an approved animal use protocol, were seropositive as determined by a whole-cell ELISA and a Western blot assay, as described previously (18). These sera were compared with sera from four adult laboratory-reared *P. leucopus* mice that were obtained from the Peromyscus Stock Center, University of South Carolina, and were seronegative as determined by the same assays.

Mouse immunization. Female, 4-week-old BALB/c mice (Jackson) were inoculated intraperitoneally with 10 μg purified protein in phosphate-buffered saline (PBS) or PBS alone emulsified with Freund's complete adjuvant and were boosted twice at 2-week intervals with the antigen solution or PBS alone in incomplete Freund's adjuvant. Plasma samples were collected before each immunization and after the final boost using the Microvette CB 300 system for capillary blood collection (Sarstedt).

Antibody reactions and assays. For experiments with arrays, human sera were diluted 1:200 in protein array blocking buffer (Whatman) that was supplemented with a lysate of *E. coli* at a final protein concentration 5 mg/ml and then were incubated at room temperature for 30 min with constant mixing (29). The arrays were rehydrated in blocking buffer for 30 min, incubated with the pretreated sera for 12 h at 4°C with constant agitation, washed in 10 mM Tris (pH 8.0)-150 mM NaCl containing 0.05% Tween 20 buffer, and then incubated with biotin-conjugated goat anti-human IgG (Fc- γ fragment-specific) serum (Jackson ImmunoResearch) that was diluted 1:200 in blocking buffer. After the array slides were washed in 10 mM Tris (pH 8.0)-150 mM NaCl, bound antibodies were detected with streptavidin conjugated with the dye PBXL-3 (Marteck). The washed and air-dried slides were scanned with a Perkin Elmer ScanArray Express HT apparatus at a wavelength of 670 nm and with an output of RGB format TIFF files that were quantitated using ProScanArray Express software (PerkinElmer) with correction for spot-specific background. When *P. leucopus* serum was used, it was diluted 1:200 in protein blocking buffer, and alkaline phosphatase-labeled goat anti-*P. leucopus* IgG antiserum (Kirkegaard and Perry) was used as the secondary antibody. Bound antibodies were detected using one-step nitroblue tetrazolium-5-bromo-4-chloro-3-indolylphosphate (BCIP) (Pierce). Arrays were scanned at 2,400 dpi (Hewlett-Packard ScanJet 8200 scanner), and after images were converted to gray scale format and inverted, they were quantitated as described above. Western blot analyses of whole lysates of *B. burgdorferi* using 10 μg protein per lane or 250 ng purified protein per lane were carried out as described previously (10). Nitrocellulose membranes were incubated with human or mouse serum at a dilution of 1:250, and bound antibodies were detected by incubation with alkaline phosphatase-labeled goat anti-human IgG antiserum or anti-mouse IgG antiserum (Jackson ImmunoResearch) at a dilution of 1:1,000. The murine monoclonal IgG antibodies to OspA (BBA15) and FlaB (BB0147) used were H5332 and H9724, respectively (9, 10).

Data analysis. The raw values from array scans were the mean average intensities of all the pixels in a pair of printed spots for each Orf or negative control; these raw values were then log transformed. (A preliminary analysis showed that there was no difference in interpretation whether or not the mean value for DNA controls was subtracted from each raw value before log transformation, and, consequently, this additional step was not included.)

The following analyses were carried out. (i) The mean and standard deviation (SD) for each Orf with all control serum samples in each panel were determined. For each Orf and for each serum sample, the number of SDs above or below the mean for the control sera in the same serum panel for the Orf in question was determined. For each sample all the Orfs that had array spots with values that were ≥ 2 or ≥ 3 SDs above the mean for the negative controls in the experiment for the given Orf were tabulated and summed. The frequencies of each Orf that appeared in this cumulative list were then determined. (ii) Bayesian microarray expression analysis and discriminatory antigen selection were performed with software adapted from Cyber-T for protein arrays (4, 86, 87). The correction of Hochberg and Benjamini was applied to control for false discoveries under the multiple test conditions (47). (iii) Cluster analyses were performed and graphic displays of array results were generated using the MultiExperiment Viewer v 4.0 software available from The Institute for Genomic Research (78). The Euclidian distance criterion with average linkage was used, and 1,000 bootstrap analyses with replacement iterations were carried out. (iv) Receiver operating characteristic curves were generated for selected sets of Orfs using the packages "e1071" and "ROCR" in the R statistical environment, available at the <http://www.R-project.org> website. (v) Standard asymptotic or exact statistical analyses of continuous data were carried out with the SYSTAT version 11 (SYSTAT Software, Inc.) software, the StatExact version 6 (Cytel Software Corporation) software, or Confidence Interval Analysis version 2.1.2 (2), available at the <http://www.som.soton.ac.uk/cia> website. Unless noted otherwise, significance tests were two sided. For means, differences, and odds ratios (OR), 95% confidence intervals are indicated below.

RESULTS

Proteome array and overall binding of antibodies. The array comprised in vitro products of 1,292 ORFs of strain B31 and an additional *ospC* allele from another strain for a total of 1,293 *B. burgdorferi* ORFs. In separate experiments array slides were incubated with samples of serum panels 1 and 2. A serum specimen from a patient with early LB in panel 1 was replicated in the same experiment. The Pearson and Spearman correlation coefficients were 0.94 and 0.87, respectively, for paired log-transformed raw intensity values; the mean \log_{10} difference between the replicates of this serum was only 0.07 (95% confidence interval, 0.06 to 0.08) for the total set of 1,293 ORFs. For the replicates of the 26 Orfs on the array, the Pearson and Spearman correlation coefficients were 0.84 and 0.84, respectively, for panel 1 and 0.81 and 0.83, respectively, for panel 2. The corresponding mean \log_{10} differences were 0.07 (95% confidence interval, 0.00 to 0.15) and 0.04 (95% confidence interval, -0.03 to 0.11) for the two panels.

The raw and log-transformed intensity values for each ORF and each panel 1 and 2 serum are shown in Tables S1 and S2 in the supplemental material, respectively. Figure 1 provides an overall summary of the data obtained by pairwise plotting of log-transformed values for each ORF for the three clinical groups of each serum panel (controls, early infection, and later infection). This figure includes the distributions for each group as a whole, as well as medians for the distributions and the results of a nonparametric analysis of the paired data. With infection and its advance, medians shifted slightly to the right in the distributions of average intensities per ORF by clinical group. More notable were the longer right-handed tails of the distributions, which are also reflected by the several outlying points in the scatter plots for patient sera versus controls and

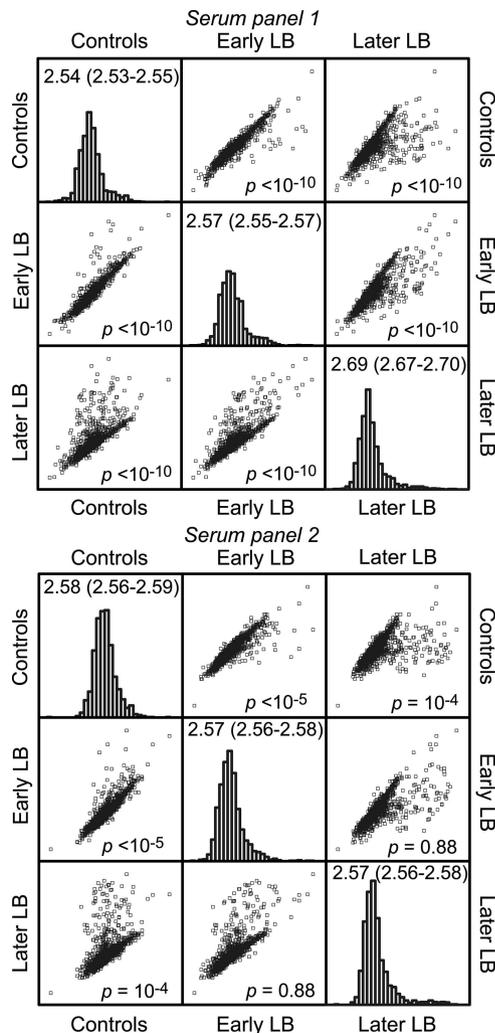


FIG. 1. Overview of genome-wide proteome array results obtained with sera from humans with LB. The graphs are two sets of frequency histograms and scatter plots of binding of antibodies in panel 1 (upper panel) and panel 2 (lower panel) sera from controls and patients with early or later LB to an array of recombinant proteins produced in vitro from a total of 1,293 *B. burgdorferi* ORFs (1,292 ORFs from strain B31 and 1 ORF from strain 297). In the frequency histograms the x axes indicate relative \log_{10} intensity values, and the y axes indicate the relative counts on an interval scale. In the scatter plots both the x and y axes indicate relative \log_{10} intensity values. The distributions in the frequency histograms are indicated along with the medians of \log_{10} intensity values with 95% confidence intervals. The scatter plots include two-sided P values obtained by using an exact Wilcoxon signed rank test. The data are shown in Tables S1 and S2 in the supplemental material.

for late-infection sera versus early-infection sera. The two panels differed in whether early-infection sera could be distinguished from late-infection sera by a nonparametric statistic (Fig. 1).

The analysis described above first averaged ORF values across clinical groups and then examined correlations for each of the 1,293 ORFs. If, instead, average ORF intensities by serum sample for all 1,293 ORFs were calculated first and then the averages by serum were compared by clinical stage, heterogeneity of the results for individual sera was observed as over-

lapping distributions for controls and patient groups. For the first serum panel the mean differences in raw intensity values between patient sera and controls were 42 (95% confidence interval, -135 to 218) for early infection and 157 (95% confidence interval, -3 to 318) for later infection; the corresponding differences from controls for panel 2 were -36 (95% confidence interval, -220 to 149) for early infection and 8 (95% confidence interval, -185 to 202) for patients with Lyme arthritis. Thus, the total amount of antibody binding to the array, which is analogous to a whole-cell assay, could not be used to assign sera to infection and control bins with confidence. More promising for this purpose was the smaller number of ORFs populating the long tail in the distributions and the “outliers” in the plots shown in Fig. 1.

By using pairwise comparisons of all sera for individual ORFs we estimated that the upper limit of the number of strain B31 Orfs that were informative as immunogens was ~ 200 of the 1,292 Orfs on the array (see the Appendix). To identify these immunogens, we used two complementary approaches. The first approach was based on an often-used criterion for setting a “cutoff” between interpretations of positive and negative for serological assays, namely, values that were ≥ 3 SDs above the average for control sera in the same run. Before using this approach, we first determined whether variances were out-of-proportion high when the mean values for control specimens increased. This would have been reflected in a significant increase in the coefficient of variation (CV) (i.e., SD divided by the mean) as the mean increased. For the 1,292 B31 Orfs and the panel 1 control sera, the mean CV was 0.115 (95% confidence interval, 0.113 to 0.117), and there was little correlation ($R = 0.06$) between the mean and the CV over the range of means. Inasmuch as the SDs for the control sera were the same in both experiments (namely, 0.85 for the 18 control sera in panel 1 and 0.84 for the 20 control sera in panel 2), normalizing the data in units of SDs allowed us to combine the data sets for the two panels. Using a simulation procedure (see the Appendix), we found that for the combined set of the later LB panel 1 and 2 sera, ORF values that exceeded the cutoff at a frequency of 6 or more times, out of a possible 39, were unlikely by chance at a one-tailed level of confidence of 0.025.

The log-normalized data for later LB sera and controls of both panels were also examined using a Bayesian statistical procedure (4), using software originally developed for DNA microarray analysis and then modified for antibody binding to proteome arrays (86, 87). For each ORF, an analysis of variance (ANOVA) comparing control sera with later LB sera was performed. In this analysis, the empirical sample variances are replaced with Bayes-regularized variance estimates. The Bayes-regularized variance is obtained by incorporating both the empirical sample variance and the variance of proteins with similar intensity levels (3, 4). This analysis produced F scores and P values that were used to rank the ORFs (see Table S4 in the supplemental material). The \log_{10} of the F score correlated with the frequencies based on a cutoff of ≥ 3 SDs ($R^2 = 0.73$ as determined by linear regression).

Identification of immunogens. We next identified the most informative of the ~ 200 immunogenic Orfs. Table 1 lists in alphabetical and numerical order the 84 Orfs whose values were ≥ 3 SDs above the control values 6 or more times out of a possible 39, had F scores of > 11 , and had corrected P values

TABLE 1. Immunogenic proteins of *B. burgdorferi* in natural infections of humans and mice

ORF ^a	Later Lyme disease (n = 39) ^b	F score ^c	P value ^d	Early Lyme disease (n = 44) ^e	Mice (n = 10) ^f	PF	Predicted to be lipoprotein ^g	Deduced gene product ^h
BB0056	12	18.7	6.0E-04	4 (5)	0		-	Phosphoglycerate kinase
BB0108	9	18.1	7.9E-04	1 (2)	0		-	Peptidylprolyl isomerase
BB0147	17	80.7	4.8E-12	5 (17)	4		-	Flagellar filament (FlaB)
BB0181	2	27.1	2.6E-05	1 (3)	0		-	Flagellar hook-associated protein (FlgK)
BB0215	10	55.3	3.3E-09	2 (7)	3		+	Phosphate ABC transporter (PstS)
BB0238	7	33.8	2.5E-06	2 (3)	0		-	Hypothetical protein
BB0260	5	19	5.4E-04	3 (4)	2		-	Hypothetical protein
BB0279	34	246.9	0	10 (16)	7		-	Flagellar protein (FlhI)
BB0283	32	149.4	0	2 (3)	1		-	Flagellar hook protein (FlgE)
BB0286	0	29.4	1.2E-05	0 (1)	7		-	Flagellar protein (FlhB)
BB0323	9	39.8	3.3E-07	1 (2)	2		+	Hypothetical protein
BB0328	2	27	2.7E-05	0 (1)	0	37	+	Oligopeptide ABC transporter (OppA-1)
BB0329	30	192.2	0	7 (9)	0	37	+	Oligopeptide ABC transporter (OppA-2)
BB0337	5	17	1.2E-03	0 (0)	0		-	Enolase
BB0348	14	59.5	1.1E-09	1 (6)	1		-	Pyruvate kinase
BB0359	15	50.3	1.3E-08	1 (2)	0		-	Carboxy-terminal protease
BB0365	15	43.8	9.9E-08	4 (8)	4		+	Lipoprotein LA7
BB0385	6	29.3	1.2E-05	2 (3)	2		+	Basic membrane protein D (BmpD)
BB0408	16	32	4.7E-06	0 (4)	4		-	Phosphotransferase system, fructose-specific IIABC
BB0476	7	19.9	3.9E-04	0 (2)	1		-	Translation elongation factor TU (Tuf)
BB0518	17	62.7	4.7E-10	3 (5)	2		-	Molecular chaperone (DnaK)
BB0543	7	45.2	6.4E-08	2 (3)	2		-	Hypothetical protein
BB0603	11	54.7	3.9E-09	3 (10)	3		-	P66 outer membrane protein
BB0649	1	20.8	2.7E-04	0 (3)	1		-	Chaperonin (GroEL)
BB0652	5	16.6	1.4E-03	2 (2)	0		-	Protein export protein (SecD)
BB0668	6	21.7	1.9E-04	1 (3)	1		-	Flagellar filament outer layer protein (FlaA)
BB0681	8	36.2	1.1E-06	1 (4)	2		-	Methyl-accepting chemotaxis protein
BB0751	6	24.6	6.5E-05	0 (1)	1		-	Hypothetical protein
BB0772	6	14.8	2.9E-03	1 (2)	0		-	Flagellar P-ring protein (FlgI)
BB0774	11	44.6	7.8E-08	1 (1)	2		-	Flagellar basal body cord protein (FlgG)
BB0805	6	13.2	5.5E-03	0 (2)	1		-	Polyribonucleotidyltransferase (PnpA)
BB0811	9	20.4	3.1E-04	2 (4)	0		-	Hypothetical protein (COG1413)
BB0844	7	43.1	1.2E-07	0 (2)	10	12	+	Hypothetical protein
BBA03	13	35.3	1.5E-06	1 (4)	3		+	Hypothetical protein
BBA04	7	12.5	7.1E-03	0 (1)	3	44	+	"S2 antigen"
BBA07	17	40.6	2.7E-07	1 (4)	0		+	Hypothetical protein
BBA15	16	28.3	1.7E-05	3 (6)	2	53	+	Outer surface protein A
BBA16	22	58	1.6E-09	1 (2)	0	53	+	Outer surface protein B
BBA19	1	22	1.7E-04	1 (3)	3	50	-	Hypothetical protein
BBA25	33	134.5	0	23 (27)	10	74	+	Decorin binding protein B
BBA34	13	76.6	1.3E-11	1 (2)	0	37	+	Oligopeptide ABC transporter (OppA-5)
BBA36	20	52.2	8.0E-09	5 (7)	7		+	Hypothetical protein
BBA40	10	17.2	1.1E-03	1 (1)	1	148	-	Hypothetical protein
BBA48	5	33.4	2.8E-06	1 (3)	0	154	-	Hypothetical protein
BBA57	9	39.9	3.3E-07	2 (6)	9		+	Hypothetical protein
BBA64	14	75.6	1.6E-11	6 (14)	7	54	+	Hypothetical protein
BBA66	7	34.4	2.0E-06	4 (6)	0	54	+	Hypothetical protein
BBB09	14	34.7	1.9E-06	0 (0)	2		+	Hypothetical protein
BBB14	13	60.9	7.3E-10	1 (1)	0		+	Hypothetical protein
BBB16	8	37.1	8.3E-07	0 (2)	0	37	+	Oligopeptide ABC transporter (OppA-4)
BBB19-A	21	79.5	6.3E-12	23 (26)	10		+	OspC type A (strain B31)
BBB19-K	24	54.1	4.5E-09	16 (21)	7		+	OspC type K (strain 297)
BBC03	5	15.1	2.6E-03	1 (4)	0	49	-	Hypothetical protein
BBC06	8	28.3	1.7E-05	1 (2)	1	95	-	EppA (BapA)
BBC10	11	15.7	2.0E-03	0 (0)	6	63	+	RevA
BBE09	4	17	1.2E-03	1 (2)	4	44	+	Hypothetical protein
BBF03	23	69.1	8.3E-11	1 (3)	0	80	+	BdrS (BdrF1)
BBF33	3	140	0	3 (21)	9		+	VlsE
BBG18	7	28.7	1.5E-05	1 (2)	0		-	Hypothetical protein
BBG33	36	286.2	0	11 (16)	10	80	-	BdrT (BdrF2)
BBH06	16	57.1	2.1E-09	2 (2)	0		+	Hypothetical protein
BBH13	30	140.7	0	5 (8)	4	80	-	BdrU (BdrF3)
BBI42	16	61.5	6.4E-10	1 (1)	0	52	+	Hypothetical protein
BBJ24	6	45.5	6.0E-08	0 (1)	0	106	-	Hypothetical protein
BBK07	13	40	3.3E-07	11 (21)	10	59	+	Hypothetical protein
BBK12	18	41.8	1.8E-07	12 (22)	9	59	+	Hypothetical protein
BBK13	8	25.5	4.7E-05	0 (3)	2	40	-	Hypothetical protein (COG2859)
BBK19	12	67.8	1.2E-10	4 (6)	8		+	Hypothetical protein
BBK23	4	19.5	4.6E-04	2 (4)	1	62	-	Hypothetical protein
BBK32	22	122.3	0	13 (17)	9		+	Fibronectin-binding protein
BBK52	3	25.4	4.9E-05	1 (3)	4	44	+	"P23"
BBK53	10	31.5	5.5E-06	0 (3)	3	52	+	Hypothetical protein
BBL03	9	23.7	9.0E-05	6 (11)	0	148	-	Hypothetical protein
BBL27	33	229.6	0	6 (11)	4	80	-	BdrO (BdrE1)
BBL39	5	23.8	8.9E-05	0 (3)	4	162	+	ErpN (CRASP-5)
BBL40	22	51.5	9.4E-09	6 (7)	10	163	+	ErpO
BBM27	18	51.6	9.3E-09	2 (5)	6	63	+	RevA
BBM34	27	153.1	0	3 (7)	6	80	-	BdrK (BdrD2)

Continued on facing page

TABLE 1—Continued

ORF ^a	Later Lyme disease (n = 39) ^b	F score ^c	P value ^d	Early Lyme disease (n = 44) ^e	Mice (n = 10) ^f	PF	Predicted to be lipoprotein ^g	Deduced gene product ^h
BBM36	3	21.5	2.1E-04	0 (2)	0	144	–	Hypothetical protein
BBN11	23	84.9	1.8E-12	2 (3)	0	152	–	Hypothetical protein
BBN27	27	143.9	0	3 (6)	5	80	–	BdrR (BdrE2)
BBN28	6	21	2.5E-04	0 (2)	0	113	+	MlpI
BBN34	31	192.1	0	8 (13)	4	80	–	BdrQ (BdrD10)
BBN38	20	56.2	2.6E-09	7 (9)	2	162	+	ErpP (CRASP-3)
BBN39	23	42.2	1.6E-07	5 (9)	9	163	+	ErpQ
BBO34	27	122.7	0	4 (6)	3	80	–	BdrM (BdrD3)
BBO39	23	71.1	5.0E-11	2 (6)	8	164	+	ErpL
BBO40	6	23.4	1.0E-04	0 (2)	6	164	+	ErpM
BBP34	31	190.9	0	6 (11)	3	80	–	BdrA (BdrD4)
BBP39	12	25.6	4.6E-05	4 (9)	9	163	+	ErpB
BBQ03	27	101.9	3.4E-14	1 (2)	5	52	+	Hypothetical protein
BBQ04	6	23.3	1.1E-04	1 (3)	4	44	+	Hypothetical protein
BBQ13	1	15.7	2.0E-03	1 (1)	2	149	–	Hypothetical protein
BBQ19	6	16.4	1.5E-03	3 (5)	2	153	–	Hypothetical protein
BBQ34	30	170.4	0	3 (7)	7	80	–	BdrW (BdrE6)
BBQ35	3	15.7	2.0E-03	1 (3)	3	113	+	MlpJ
BBQ40	6	11.7	9.8E-03	3 (5)	0	32	–	Partition protein
BBQ42	30	179.6	0	6 (9)	1	80	–	BdrV (BdrD5)
BBR12	6	13.3	5.3E-03	0 (2)	1	153	–	Hypothetical protein
BBR35	8	30.1	9.2E-06	1 (3)	0	80	–	BdrG
BBR42	14	47.3	3.4E-08	1 (1)	5	164	+	ErpY
BBS30	0	20.5	3.0E-04	0 (4)	4	113	+	MlpC
BBS41	18	29.8	1.0E-05	1 (3)	7	164	+	ErpG

^a Bold type indicates an ORF that had a P value of <0.005 but whose frequency for later Lyme disease sera was <6.
^b The numbers are the numbers of serum samples whose values were ≥3 SDs above the mean of the controls for the panel. n is the number of individuals in the group for combined panel 1 and 2 sera.
^c The F score is the Bayes-regularized variance (see the text).
^d The P value is the corrected P value (0, P < 1.0E-14).
^e The numbers are the numbers of LB patient serum samples whose values were ≥3 SDs or ≥2 SDs above mean of the controls for panels 1 and 2. n is the number of individuals in the group for combined panel 1 and 2 sera.
^f The numbers are the numbers of *P. leucopus* sera (out of 10) whose values are ≥3 SDs above the mean for four control *P. leucopus* mice.
^g +, protein predicted to be a lipoprotein; –, protein not predicted to be a lipoprotein.
^h Alternative protein designations are given in parentheses.

of <0.001. The Orfs with the highest frequencies of values that were ≥3 SDs above the control values were BBG33, BB0279, BBL27, and BBA25. An additional 19 Orfs, for a total of 103 (8.0%) out of 1,292 Orfs, had P values of <5 × 10⁻⁴ as determined by the Bayes-regularized analysis. This additional group included VlsE (BBF33) and the chaperonin GroEL (BB0649). The mean numbers of amino acids were 293 (95% confidence interval, 266 to 320) for the 103 Orfs on the list, compared to 260 (95% confidence interval, 248 to 272) for the other 1,190 Orfs (P = 0.11, t test). Thus, immunogenicity was not associated with length of the protein. Moreover, there was no difference between the two groups of Orfs in terms of the amount of protein on the array, as measured by the raw values for antibody binding to the hemagglutinin moiety of the recombinant proteins; the values were 2,816 (95% confidence interval, 2,369 to 3,349) for the 103 Orfs and 2,618 (95% confidence interval, 2,493 to 2,750) for the other Orfs (P = 0.42).

Several proteins that were known antigens and valuable for serodiagnosis were on the list. These proteins included FlaB (BB0147) (9, 45), the P66 outer membrane protein (BB0603) (5, 16), OspA and OspB (BBA15 and BBA16) (48), decorin-binding protein B (BBA25) (37, 46), OspC (BBB19) (68, 96), fibronectin-binding protein (BBK32) (71), and VlsE (BBF33) (54, 56). The other reactive Orfs that were previously reported to elicit antibodies during infections of humans or experimental animals were as follows: LA7 (BB0365) (53, 94), the chaperonins DnaK (BB0518) and GroEL (BB0649) (58), FlgE

(BB0283) (51), some Erp proteins (59, 85), oligopeptide ABC transporters (OppA; BB0328, BB0329, BBA34, and BBB16) (25, 28, 65), “S2 antigen” (BBA04) (36), the paralogous BBA64 and BBA66 proteins (65), RevA proteins (BBC10 and BBM27) (41, 65), EppA/BapA (BBC06) (63), Mlp proteins (BBN28, BBQ35, and BBS30) (70), and some Bdr proteins (99).

There were several Orfs that previously either were not recognized as immunogens during infection or had received little attention. Notable among this group were the following: (i) the paralogous BBK07 and BBK12 lipoproteins; (ii) BBK19 and BBK53, two other lipoproteins encoded by plasmid lp36; (iii) several more flagellar apparatus proteins, including FliL (BB0279), FlaA (BB0668), and FlgG (BB0774); (iv) additional paralogous family (PF) 44 proteins (BBE09, BBK53, and BBQ04); (v) BB0260, BB0323, BB0543, and BB0751, hypothetical proteins encoded on the chromosome; (vi) BBA03, BBA07, BBA36, and BBA57, hypothetical proteins or lipoproteins uniquely encoded by lp54; and (vii) BBG18 and BBH06, unique hypothetical proteins encoded by other plasmids. On the list of new immunogens there were only a few chromosome-encoded Orfs that were homologous to proteins having established functions in other bacteria, such as the phosphate ABC transporter PstS (BB0215), pyruvate kinase (BB0348), a carboxy-terminal protease (BB0359), and a methyl-accepting chemotaxis protein (BB0681).

Whereas plasmid-encoded Orfs accounted for 536 (41%) of the 1,292 B31 Orfs on the array, 70 (69%) of the 102 immu-

nogenic Orfs of strain B31 are plasmid encoded (OR, 3.1 [95% confidence interval, 2.0 to 4.9]; exact $P < 10^{-6}$). Fifty-nine (58%) Orfs, all but two of which were plasmid encoded, belonged to 1 of 26 PFs. Of a possible 174 Orfs that belong to 1 of these 26 PFs, 114 (66%) were included as amplicons on the array. The greatest representation was that of PF 80, which comprises the Bdr proteins; 12 (92%) of a possible 13 Orfs were on the list of 83 Orfs. These Orfs included high-ranking BBG33 and BBL27 proteins. Other PFs with three or more representatives on the list were the PFs containing the Erp proteins (PFs 162 to 164), oligopeptide ABC transporters (PF 37), Mlp proteins (PF 113), the "S2 antigen" and related proteins (PF 44), and a set of hypothetical proteins with unknown functions (PF 52).

For tabulation of the plasmid locations of the ORFs shown in Table 1, we excluded pseudogenes and ORFs that were less than 300 nucleotides long (21). The sizes of linear plasmids lp38 (38,829 nucleotides) and lp36 (36,849 nucleotides) are similar. Only 1 of lp38's 17 ORFs, BBJ24, was among the ORFs encoding high-ranking antigens, but 8 of the 19 lp36 ORFs were (OR, 11.6 [95% confidence interval, 1.2 to 548]; $P = 0.03$). The presence of plasmid lp36 has been associated in one study with infectivity or virulence in a mammalian host (49), as has been the presence of lp25 in another study (72), but only BBE09 of the 10 ORFs of lp25 were among the ORFs encoding immunogens.

Forty-eight (47%) of the 102 immunogens of strain B31 are lipoproteins as determined by prediction or empirical documentation. Of the 756 chromosome-encoded Orfs included in the array, only 32 (4%) are lipoproteins, but, as shown in Table 1, 7 (21%) of the 33 chromosome-encoded Orfs among the immunogens are lipoproteins (OR, 6.1 [95% confidence interval, 2.1 to 15.8]; $P = 0.001$). Whereas 85 (16%) of the 536 plasmid-encoded Orfs on the array are predicted lipoproteins, 41 (59%) of the 70 plasmid-encoded proteins of strain B31 on the antigen list are predicted lipoproteins (OR, 7.6 [95% confidence interval, 4.3 to 13.4]; $P < 10^{-12}$). In addition to five documented outer membrane proteins (OspA, OspB, OspC, VlsE, and P66), the following three hypothetical proteins among the immunogens were predicted to localize to the outer membrane by the PSORT algorithm for double-membrane bacteria (40): BB0260, BB0751, and BB0811.

Stage of infection. In general, sera from early in infection reacted with fewer antigens per serum sample and antigens from a narrower list of antigens. For 20 (83%) of the 24 panel 1 early LB cases there was at least one Orf in Table 1 whose value exceeded the 3-SD cutoff. Of the four cases for which there was not at least one Orf whose value exceeded the 3-SD cutoff, three (75%) were seronegative as determined by ELISA and IgG and IgM Western blotting. Of the 20 cases of early infection for serum panel 2, 17 (85%) had at least one Orf whose value was ≥ 3 SDs. For the 37 samples with one or more reactive Orfs, the number of Orfs whose values were above the threshold ranged from 1 to 37, and the median number was five Orfs per sample. For the 84 antigens identified by the first analysis, the values for 69 (82%) were above the threshold for at least one of the early-infection sera (Table 1). In most cases, the following 15 Orfs whose values fell below the cutoff with all early sera were also among the least prevalent Orfs for sera obtained later in disease: BB0408, BB0476, BB0751, BB0805,

BB0844, BBA04, BBB09, BBB16, BBC10, BBJ24, BBK13, BBK53, BBN28, BBO40, and BBR12. The Orfs whose values exceeded the cutoff in at least 10 of the 37 samples were, in descending order, BBA25 (DbpB), BBB19 (OspC type A), BBB19 (OspC type K), BBK32 (fibronectin-binding protein), BBK12, BBG33 (BdrT), BBK07, and BB0279 (FlilL).

Sera of the 10 patients with refractory Lyme arthritis were compared with sera of the 10 patients with arthritis that responded to antibiotic therapy. As determined by a *t* test and nonparametric rank test of log-transformed values, there was not a significant difference ($P > 0.05$) between the two groups for any of the 1,293 Orfs, including both OspC proteins.

White-footed mouse antibodies. Using the same batch of genome-wide arrays, we examined the reactions of sera from 10 *P. leucopus* mice that were captured in an area in which the level of *B. burgdorferi* infection of mice approached 100% by the end of the transmission season (18). All 10 mice were seropositive as determined by the whole-cell assay and Western blot analysis (18). These sera were compared with sera from four laboratory-reared *P. leucopus* mice. As described above, we calculated the number of SDs above or below the mean of the controls for each Orf and each mouse serum. Of the 103 Orfs shown in Table 1, only 30 (29%) were not represented at least once among the Orfs with values of ≥ 3 SDs with *P. leucopus* sera. The highest frequencies (≥ 7 of 10 sera) were those of the following Orfs, in alphabetical order: BB0279, BB0286, BB0844, BBA25, BBA36, BBA57, BBA64, BBB19 (OspC types A and K), BBF33, BBG33, BBK07, BBK12, BBK19, BBK32, BBL40, BBN39, BBO39, BBP39, BBQ34, and BBS41. Thirteen Orfs had frequencies of ≥ 5 among the 10 *P. leucopus* sera but were not among the high-ranking Orfs with human sera (Table 1). These Orfs included two hypothetical proteins (BB0039 and BB0428), two members of PF 143 (BBP26 and BBS26), and the BBK50 protein, another lp36-encoded protein. But also represented among the high-ranking Orfs with *P. leucopus* sera were members of PFs, at least one of which was frequently recognized by human antibodies, including two additional PF 113 proteins, MlpH (BBL28) and MlpA (BBP28); another PF 164 protein, ErpK (BBM38); and an additional PF 54 protein, BBA73. Overall, there was considerable overlap in the sets of immunogens for humans and *P. leucopus* infected with *B. burgdorferi*.

Second array. To confirm the results described above, we produced a second array with 66 recombinant proteins selected from the 103 Orfs shown in Table 1. The second array contained three additional proteins that were not cloned for the first array. Two of these, BB0383 (BmpA or P39 protein) and BB0744 (P83/100 protein), are among the 10 signal antigens for a commonly used criterion for Western blot interpretation (33). The third additional ORF was BBA24 or decorin-binding protein A (DbpA). The smaller arrays were incubated with 12 later LB sera and three control sera from panel 1. Figure 2 shows the results in a two-color gradient format with an accompanying cluster analysis. BB0383 and BB0744 clustered with several other proteins that were frequently bound by antibodies of LB sera, including FlaB (BB0147), BB0279 (FlilL), VlsE, and DbpB (BBA25). The patterns of reactivity with different patient and control sera were essentially the same for these antigens. This was demonstrated by correlations between the Orfs; for BB0383, the R^2 values for BB0279, BB0147, VlsE,

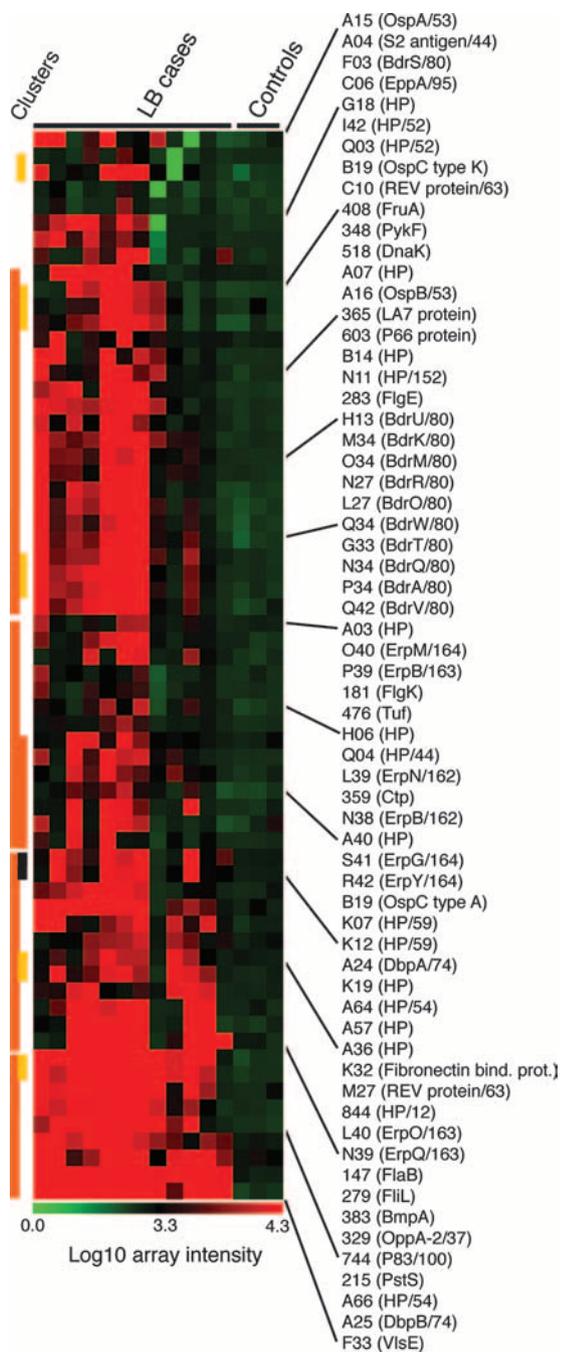


FIG. 2. Two-color display of a Euclidian distance cluster analysis of a small array with immunogenic Orfs of *B. burgdorferi*. The array was incubated with 12 sera from individuals with later LB (panel 1) or three control sera. In addition to selected Orfs from Table 1 the following three proteins were used: BB0383 (BmpA), BB0744 (P83/100), and BBA24 (DbpA). The Orf designations do not include “BB.” Protein names and hypothetical proteins (HP), as well as the PFs, are indicated on the right. Clusters are indicated on the left. The levels of bootstrap support (1,000 iterations) are as follows: orange, >50%; yellow, >60%; and black, 100%. A scale for log₁₀ intensity values is at the bottom.

and BBA25 were 0.90, 0.86, 0.70, and 0.58, respectively, and for BB0744, the corresponding R^2 values were 0.91, 0.81, 0.71, and 0.74, respectively. DbpA (BBA24), whose sequence is genetically more diverse than that of DbpB across strains (76), was

less frequently reactive with the collection of patient sera than DbpB. Thus, addition of the P83/100, BmpA, or DbpA protein to the array provided little or no additional discriminatory power.

Figure 2 also shows clustering of the Bdr proteins and the BBK07/BBK12 proteins but not of the two OspC proteins in terms of their patterns and intensities of reactivities with this set of sera. The relationship between clustering of Orfs and amino acid sequence identity was examined by plotting normalized values for individual panel 1 control, early LB, and later LB sera and for selected pairs of Orfs (Fig. 3). There was high correlation between the paralogous proteins BBK07 and BBK12 and between two Bdr proteins, BBG33 and BBL27, in the serum antibody reactions. The data corresponded to amino acid sequence identities of 87 and 80%, respectively. In contrast, there were greater differences in the reactivities of sera with the two OspC proteins, even though the overall level of identity between them was close to that of the two Bdr proteins. Lower still was the correlation between two high-ranking proteins which are similar sizes but are not homologous, BBK07 and BBA25.

Purified proteins. We selected five Orfs for further investigation as purified recombinant proteins: BB0279 (FlilL), BB0283 (FlgE), BBA25 (DbpB), BBG33 (BdrT), and BBK12. Western blot analyses were carried out with sera from 17 patients with later LB and five panel 1 controls (Fig. 4). Binding that was noted in the array was confirmed by Western blotting; no bands were observed with the control sera. We then used different amounts of proteins BBA25, BBG33, BBK12, and BB283 over a 30-fold range in an array format and incubated the chips with the same patient and control sera (Fig. 5). While for some proteins the binding by control sera increased with higher protein concentrations, the log-transformed raw values for patient sera changed little over the concentration range used, an indication that the absolute amount of protein in the spots of the high-throughput array over this range was not a major determinant of the amount of antibody binding as estimated by digitization of the signals for this study. When binding to in vitro-produced proteins on the array was compared to binding to different amounts of a purified protein for a given Orf for a standard curve, we estimated that the amounts of Orfs in the genome-wide array were 50 to 400 pg protein per spot.

While detection of antibody to an Orf was evidence of expression of the predicted polypeptide, this evidence was indirect. We used one of the purified proteins, BBK12, to immunize mice and thereby provide a reagent for more direct documentation of expression. This Orf was chosen because it and the product of a paralogous gene, BBK07, had not been previously reported to be immunogenic. In fact, there was little previous comment on either of these proteins beyond their annotation as hypothetical lipoproteins with unknown functions. Figure 6 shows a Western blot in which lysates of low-passage or high-passage strain B31 were incubated with the anti-BBK12 antiserum or monoclonal antibodies to OspA (BBA15) or FlaB (BB0147). (Because of the probable antigenic cross-reactivity between BBK07 and BBK12, as the analysis shown in Fig. 3 suggests, we could not assume that the antiserum could easily distinguish between the two Orfs.) As expected, FlaB and OspA were expressed by both low- and

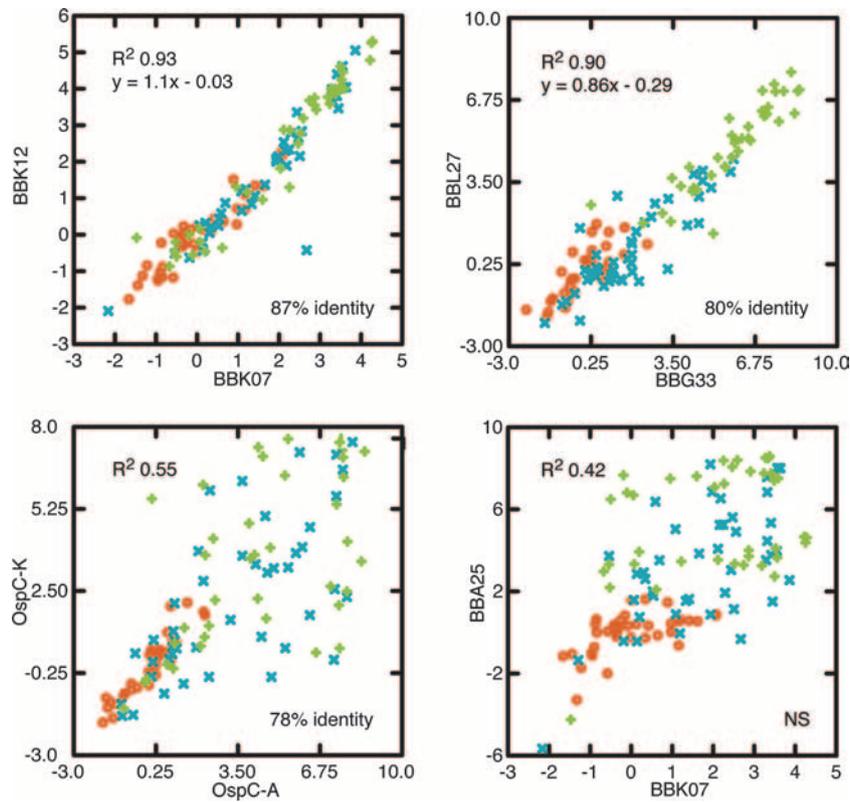


FIG. 3. Scatter plots of array intensity values normalized in units of SDs above or below the mean for the controls of each panel. Each plot shows values for pairs of selected Orfs reacted with sera of controls (red circles) and patients with early LB (blue multiplication signs) or later LB (green plus signs) of panels 1 and 2. The coefficients of determination (R^2) for all plots, as well as the linear regression equations for the upper two plots, are shown. The levels of identity of aligned amino acid sequences of the three pairs of homologous proteins are indicated; BBK07 and BBA25 are not significantly (NS) similar.

high-passage isolates of strain B31. In contrast, the BBK12 and/or BBK07 protein was detected in low-passage cells but not in high-passage cells. This experiment not only confirmed that there was expression of either BBK12 or BBK07 or both but also showed that loss of expression of these proteins was associated with high passage. Thus, one explanation for the previous lack of recognition of informative antigens, such as BBK07 and BBK12, was that higher-passage cell populations, which were often used as the basis of diagnostic assays, did not

express the proteins, either because of plasmid loss or because of transcriptional or translational failure.

How many antigens are sufficient? While our goal was not to establish a final set of antigens that in combination were specific for *B. burgdorferi* infection, this study did permit estimation of the minimum number of antigens that would be needed to achieve this goal. For this, we studied the discriminatory power of different sets of ORFs using receiver operating characteristic (ROC) curves, where the false-positive rate (1 – specificity) is the x axis and the true positive rate (sensitivity) is the y axis for different thresholds of the underlying classifier. The area under the curve (AUC) summarizes the results. An AUC of 1.0 indicates a perfect classifier, while an AUC of 0.51 (95% confidence interval, 0.38 to 0.64) is the expected value for a classifier that works by chance for the data set, as inferred by the method of Truchon and Bayly (89). The log-transformed data for controls and later LB sera from both panels were used for this analysis. First, ROC curves were generated for single antigens to assess the ability to separate the control and disease. The Orf number is the rank based on the Bayes-regularized ANOVA F score (see Table S4 in the supplemental material). The top Orfs discriminate very well. The first nine Orfs all have an AUC of >0.95 , and further down the rank, the ability diminishes. The 25th immunogen has an AUC of 0.90, the 50th immunogen has an AUC of 0.85, the 100th immuno-

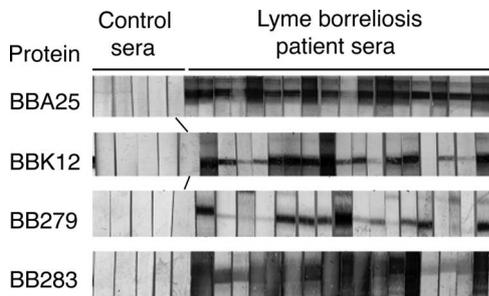


FIG. 4. Western blot analysis of purified recombinant proteins encoded by ORFs BBA25 (DbpB), BBK12, BB279 (FlhL), and BB283 (FlgE) incubated with sera of 17 patients with later LB or five panel 1 controls. Binding of antibody was detected with alkaline phosphatase-labeled secondary antibody to human IgG as described in the text.

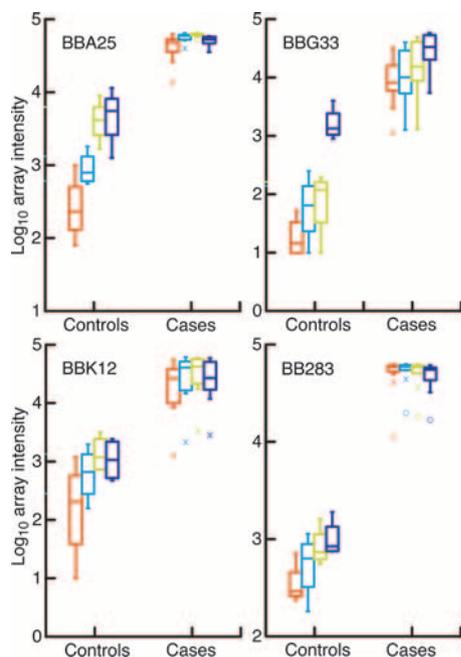


FIG. 5. Binding of antibodies in human LB sera to purified proteins on arrays. The plots are box-whisker plots of log-transformed intensity values for the binding of panel 1 sera from patients with later LB ($n = 17$) or controls ($n = 5$) with purified recombinant proteins encoded by ORFs BBA25, BBG33, BBK12, and BB283 at concentrations of 0.03 mg/ml (red), 0.1 mg/ml (light blue), 0.3 (yellow), and 0.9 mg/ml (dark blue). Each box indicates the first and third quartiles, and the line inside the box is the median. The $1.5\times$ interquartile range is indicated by the vertical line bisecting the box, and values outside this range are indicated by asterisks and by circles.

gen has an AUC of 0.74, and the 165th Orf has an AUC of 0.65, which still exceeds the upper 95% confidence interval for random expectations for the AUC.

To extend the analysis to combinations of antigens, we used kernel methods and support vector machines, as described by Vapnik (92), to build linear and nonlinear classifiers. Different kernels, including linear, polynomial, and radial basis function, were evaluated. Only the radial basis function kernel showed an increase in the AUC when noise was added, and accordingly, this kernel was chosen for subsequent simulations in which noise was introduced. For each data set, the support vector machines were tuned using a wide parameter sweep to achieve the best gamma and cost values. Results were validated with 10 runs of threefold cross-validation. As input to the classifier, we used the highest-ranking 2, 5, 25, and 45 Orfs on the basis of either Bayes-regularized ANOVA F scores or frequencies of later LB sera exceeding a 3-SD cutoff. The results of two ranking schemes were similar, and only the frequency ranking results are shown in Fig. 7.

For the present data set, there were negligible differences in the ROC curves obtained using 2, 5, 25, or 45 antigens. The mean AUC values over the 10 validation runs were >0.98 for two antigens and a perfect 1.0 for five or more antigens. The unsurpassable performance in this experiment with relatively few antigens can be attributed to the high discrimination provided by the first several antigens on the list by themselves. In a realistic diagnostic setting with sera coming from various

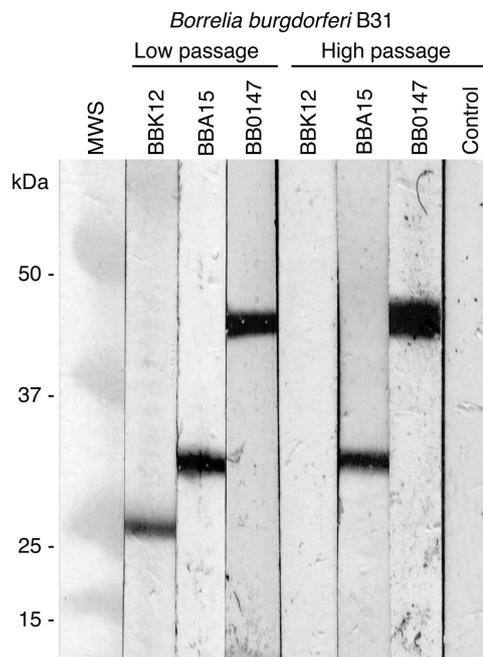


FIG. 6. Western blot analysis of whole-cell lysates of low-passage and high-passage *B. burgdorferi* strain B31 with mouse antiserum to recombinant BBK12 or with murine monoclonal antibodies to BBA15 (OspA) or BB0147 (FlaB).

sources and backgrounds and with interoperator variances, one would expect some addition of noise in the data. To further examine how combinations of antigens increase the discriminatory power, we explored two different noise models and their effects on the classifiers. The noise model involves the addition

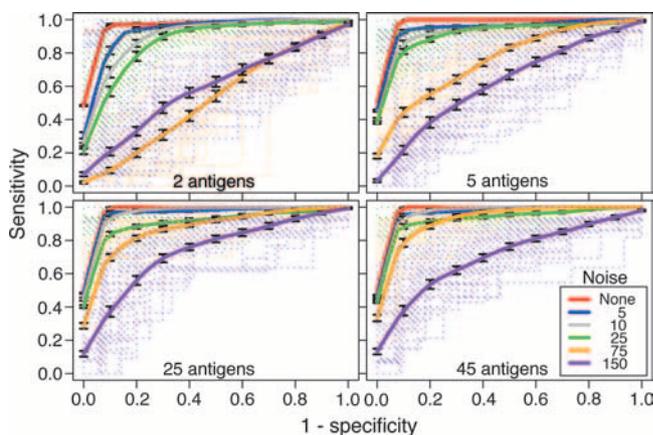


FIG. 7. Estimation of the number of immunogens for assays with high sensitivity and specificity. The graphs show four receiver operating characteristic curves for nonlinear classifiers with different sets of Orfs and the effect of increasing the amounts of uniform Gaussian noise with a mean of 0 and an SD of 5, 10, 25, 75, or 150. The antigens in sets containing 2, 5, 25, and 45 antigens were selected in order of their ranking by the Bayes-regularized analysis (see Table S4 in the supplemental material). The solid lines indicate the average with standard error over cross-validation runs calculated at stepped ($1 - \text{specificity}$) points. The error bars indicate 95% confidence intervals. The dotted lines indicate the performance for each of 10 threefold cross-validation iterations.

of uniform Gaussian noise. Each point (u) in the data set has some noise added such that $u' = u + N(\mu = 0, \sigma^2 = s)$, where s is constant across the whole data set. Noise levels are generated by scaling s . In general, using more antigens in the classifier increases resistance of the simulated assay to noise. All of the classifiers discriminate very well with low noise levels. For the two-antigen classifier, the AUC dropped to the value expected by chance by the time noise was at a scale of 75. The five-antigen classifier value dropped to 0.6 with a noise level of 150. The 25- and 45-antigen classifiers still performed relatively well, with mean AUC values of 0.74 and 0.71, respectively. Hence, based on the criteria of high predictive value and robustness in the face of increasing noise, 25 antigens were as informative as 45 antigens.

DISCUSSION

The genome-wide protein array for *B. burgdorferi* allowed comparison of far more proteins than could be compared previously with one-dimensional Western blots (8, 24, 33). While comparable numbers of proteins for analysis might theoretically be obtained with two-dimensional electrophoresis (66), scarce immunogens in the lysates would be overlooked. Moreover, unless the microbe's cells were taken directly from an infected animal, informative antigens that were expressed only in vivo would be not be included from samples subjected to electrophoresis. Our study of natural infections of humans and white-footed mice with *B. burgdorferi* followed genome-wide array analyses of antibody responses to poxvirus infections in humans immunized with a smallpox vaccine and to *F. tularensis* infections in experimental animals (29, 35, 87) and ELISA format studies of *T. pallidum* ORFs (11, 61). The major emphasis of the previous studies was identification of immunogens after immunization with whole microbes or during infection. We had the same goal for our study of natural infections that occurred in two very different ecological settings: (i) patients with different stages of Lyme disease, including the arthritis in later disease, and (ii) white-footed mice, which are a major reservoir host of *B. burgdorferi* in the United States and in which infection is nearly universal in enzootic areas. As discussed below, the goal of discovery of new antigens was met: several new immunogens were identified among the Orfs of *B. burgdorferi*.

Of equal interest was a second question: how many of the predicted proteins of this pathogen elicit an antibody response during natural infection? For this, we were concerned with the set of proteins that were not demonstrably immunogenic. Only by including most of a genome's ORFs in the experiment could one address this question, which as a general principle is relevant to many other infectious diseases. Important for hypothesis testing for this second goal was the likelihood of false negatives or type II errors. If minimizing false positives or type I errors (i.e., inaccurately identifying an Orf as an immunogen) was the experimental design challenge for the first goal, then minimizing false negatives (i.e., overlooking Orfs that were truly immunogenic) was the challenge for the second goal. In the present study, type II errors could happen for several reasons.

Indisputably, failure to amplify, clone, and then express a given ORF would lead us to miss an Orf that was actually

immunogenic. Of the ~20% of the Orfs that were absent from the array, undoubtedly some elicit antibody responses during infection. But in many of these cases, the missing Orf was a member of a PF, at least one member of which was represented in the array. Other ORFs were not included because they had characteristics of pseudogenes. Taking these considerations into account, we estimated that at least 90% of the nonredundant ORFs that were true genes were included in the array analysis. When called for, we successfully amplified some missing ORFs in reattempts using either the original primers or modifications of the primers. In these instances, addition of the antigens missing from first array to a second array did not materially change the results. This suggests that returns diminish as further efforts to fully constitute the array consume greater resources.

Another basis for type II errors would be posttranslational modifications that are important for antibody recognition that occur in *B. burgdorferi* but not in *E. coli*. While we cannot rule out a limitation to the study for this reason, there is no evidence or only scant evidence that glycosylation or a similar posttranslational modification affects antigenicity in *Borrelia* spp. The most prevalent protein modification in *Borrelia* spp. appears to be the addition of a lipid moiety to the N terminus of the processed proteins in a fashion typical of many types of bacteria. While *E. coli* cells are capable of carrying out this lipidation function for recombinant *Borrelia* proteins, this activity did not occur in the acellular transcription-translation reactions used here. This indicates that the significantly greater representation of lipoproteins among immunogens than that expected based on a lipoprotein's size among all Orfs was not attributable to antibodies to the lipid moieties themselves. Instead, the comparatively greater immunogenicity of lipoproteins may be a consequence of the mitogenicity and adjuvant-like qualities of bacterial lipopeptides.

For the 1,292 B31 ORFs that were successfully amplified, cloned, and expressed, some of the products may have been overlooked as immunogens because their epitopes are conformational and proper folding was not achieved in the in vitro reaction or subsequently when the polypeptide was printed. We cannot rule this possibility out. But the correct calls for the well-established antigens included in the array, such as OspC, FlaB, P66, P83/100, BmpA, fibronectin-binding protein, and VlsE, among others, as "immunogens" indicate that there were few instances of type II errors on the basis of loss of conformational epitopes or some other artifact of the procedures.

Another limitation of the study, at least in the case of the human sera, was the restriction of secondary antibodies to antibodies that were specific for IgG. By failing to account for IgM antibody binding, we may have underestimated the total number instances in which the Orfs were recognized by antibodies during early infection. However, we do not suspect that this effect was great if it occurred at all. There was no instance of an Orf that was recognized by antibodies in sera from early infection and not by antibodies obtained later in the disease. Our rationale for limiting antibody detection to IgG was the generally poorer specificity of IgM-based assays for Lyme disease (34, 91). We recognize the importance of eventually evaluating antigens for their predictive value with IgM as well as IgG antibodies, but the focus here was on identification of

immunogens with the greatest informative value (that is, with high specificity as well as sensitivity).

Notwithstanding the actual and theoretical limitations of the study, we concluded that the array results were not confined to identification of new immunogens but could also be used to gauge the proportion of proteins that are not immunogens. As far as we know, this perspective on immune responses during natural infection is unique among studies of proteomes of bacteria, fungi, or parasites. By taking this perspective, we estimated that the number of Orfs that elicited antibodies in at least some individuals that were infected was about 200, or ~15% of the 1,292 Orfs subjected to analysis with two panels of sera representing different stages of infection. Three types of data supported this conclusion: the magnitude of sign differences between pairs of LB patient sera with control sera (see Table A1 in the Appendix), the number of Orfs with corrected, regularized *P* values < 0.01, and number of Orfs with areas under the ROC curve that exceeded the upper confidence limit for random expectations (see Table S4 in the supplemental material). Of this larger set of immunogens, ~100 were broadly enough reactive across several LB serum samples that they could be used to distinguish groups of infected individuals from groups of controls. This interpretation also seemed to hold true for white-footed mice, which generally recognized the same subset of proteins as humans. The absolute number of distinct (i.e., non-cross-reactive) antigens is probably less than the first accounting suggested, because of the heavy representation of proteins in PFs on the immunogen list (Table 1). The several Bdr proteins on the list could probably be replaced in an array by one or two Bdr proteins with no loss of sensitivity.

The question of the minimal set of antigens necessary for discrimination between sera from patients and sera from controls was also addressed by the ROC curve analysis (Fig. 7). The introduction of increasing levels of noise provided a rough simulation of applying an assay in practice (that is, in different locations, at different times, and with different operators). It also allowed assessment of the effect of different amounts of heterogeneity in the total population. By this measure, 25 antigens provided more robustness in the face of increasing noise than 2 or 5 antigens, while expansion of the set to 45 antigens provided marginal if any advantage over 25 antigens. Whether the minimal set can be further reduced within the range from 5 to 25 antigens remains to be determined by testing the various criteria with new serum panels.

To sum up, we estimated that proteins that detectably elicit antibodies during natural infection constitute about 15% of the polypeptides that might be expressed. Incorporation of $\leq 2\%$ of the total Orfs in an assay appears to be sufficient to provide high levels of sensitivity and specificity. Our attention now turns to what the high-value immunogens are. In the course of this study, we discovered several protein antigens of *B. burgdorferi* that have promise for serodiagnosis of LB but which were unappreciated as immunogens during infection. These previously unknown antigens appear to be as informative as other proteins, such as FlaB, OspC, P66, BmpA, and VlsE, that have established value for LB serodiagnosis. In addition, in this study we also “rediscovered” several other proteins that may have been observed in a limited number of studies to be immunogenic in either natural or experimental infections but

whose value had not been confirmed or which had not been further developed. Among these are the Bdr proteins.

We compared the list of immunogenic proteins identified by proteome array analysis with lists of genes that were more highly expressed under various conditions simulating infection in the natural hosts and were reported by Revel et al. (74), Ojaimi et al. (67), Brooks et al. (13), and Tokarz et al. (88). The concurrence between the proteome list and the four DNA array lists was greatest for the study of Revel et al., and accordingly, this study was the study used for comparison. Revel et al. employed three experimental conditions: (i) 23°C and pH 7.4 in broth medium, which represented the environment in the unfed tick; (ii) 37°C and pH 6.8 in broth medium, which represented the environment in ticks as they are feeding on a host and transmitting *B. burgdorferi*; and (iii) a dialysis chamber in the peritoneum of rats. Of the 79 Orfs that showed a ≥ 2 -fold increase in mRNA under fed-tick conditions in comparison to unfed-tick conditions, the following 23 (29%) were among the high-ranking immunogens: BB0323, BB0329, BB0365, BB0668, BB0681, BB0844, BBA03, BBA07, BBA25, BBA34, BBA36, BBA66, BBB19, BBI42, BBK07, BBK13, BBK32, BBK53, BBL40, BBM27, BBO40, BBP39, and BBQ03. Four of these Orfs are encoded by the lp36 plasmid. Among the 19 Orfs whose expression was found by Revel et al. to significantly increase in dialysis chambers in comparison to conditions mimicking unfed ticks, 5 (26%) were on the antigen list. The only three Orfs whose expression decreased under conditions associated with mammalian infection were BBA15 (OspA) and BBA16 (OspB), whose expression was known to decrease in the fed ticks and during early infections in mammals (32, 80, 81), and BB0385 (BmpD). Thus, there was an association between the upregulation of genes in the fed ticks and mammals and the immunogenicity of the gene products in infected humans.

Western blots of two-dimensional electrophoresis gels provide greater resolution than one-dimensional gels and allow detection of less abundant immunogens in lysates. Nowalk et al. performed such a proteomic analysis with the same samples that constituted serum panel 1 (66). Fifteen of the 21 proteins identified by Nowalk et al. as immunogens were also high-ranking Orfs in the present study. These proteins include four Erp proteins (BBL39, BBL40, BBN38, and BBP39), three oligopeptide ABC transporters belonging to PF 37 (BB0328, BB0329, and BBB16), two PF 54 proteins (BBA64 and BBA66), a RevA protein (BBM27), and the unique hypothetical protein BBA03, as well as the established antigens BB0147 (FlaB), BB0365 (LA7), BB0603 (P66), and BBA15 (OspA).

The large number of proteins newly identified as immunogenic precludes discussion of each of them in depth here. Instead, we limit our remarks to the Bdr proteins (PF 113), flagellar apparatus proteins, and BBK07 and BBK12, the two members of PF 59. Of all the PF proteins, the Bdr proteins were the most prevalent among the Orfs shown in Table 1. We previously reported that LB patients, but not controls, had antibodies to some of the Bdr proteins (99), but in that study we did not include BdrT (BBG33), the highest-ranked Bdr protein here. While proteins in PFs tend in general to be more immunogenic than other non-PF Orfs, if only because of their multiple versions in a cell, the Bdr proteins may be doubly immunogenic because they have intramolecular repeats as well

TABLE A1. Pairwise comparisons of reactivities of sera with proteome array of *B. burgdorferi*

Reference clinical group	<i>n</i>	Controls		Early LB		Later LB	
		Mean (95% CI) differences in sign between serum pairs ^a	<i>P</i> ^b	Mean (95% CI) differences in sign between serum pairs ^a	<i>P</i> ^b	Mean (95% CI) differences in sign between serum pairs ^a	<i>P</i> ^b
Controls	18			-7.1 (-32 to 18)	0.5	-188 (-217 to -159)	<10 ⁻⁵
Early LB	24	-7 (-37 to 23)	0.9			-198 (-213 to -184)	<10 ⁻⁶
Later LB	19	106 (36 to 177)	<10 ⁻⁵	109 (35 to 184)	<10 ⁻⁵		

^a The sera were panel 1 sera from controls and from patients with early and later LB. CI, confidence interval.

^b Determined by the exact Wilcoxon signed-rank test.

(98). The number of copies of the peptide TKIDWVEKN LQKD or a variation of this peptide in a Bdr sequence determines the size of the protein. The BBG33 protein, which is 266 amino acids long, is the largest Bdr protein encoded by the B31 genome. Most of the other Bdr proteins are less than 200 residues long. If the internal repeats are immunodominant epitopes, then BdrT would display more of these repeats for the binding of antibodies than other Bdr proteins and, consequently, generate higher spot intensities. The coefficient for BBL27 regressed on BBG33 is 0.86, and the *y* intercept is -0.29 (Fig. 3), an indication of lower levels of binding across all sera to the shorter Bdr. BdrT, also called BdrF2, has been reported to be upregulated in "host-adapted" *B. burgdorferi* and to be specifically expressed during early infection in mice (75, 97).

This study revealed that several flagellar apparatus proteins besides FlaB flagellin (BB0147), the FlgE hook protein (BB0283), and the FlaA protein (BB0668) (42, 51, 69, 77) elicit antibody responses during infection. Brinkmann et al. found that FlgE of *T. pallidum* was frequently bound by antibodies in sera from patients with syphilis (11). FliL (BB0279) stood out among this larger group of flagellar antigens because of the frequency with which it was recognized by both human and white-footed mouse serum. Indeed, the field mice had antibody to FliL more frequently than they had antibody to FlaB, the long-standing flagellar antigen of choice for diagnosis. FliL has 178 residues and is the flagellar basal body-associated protein, which as an inner membrane protein interacts with the cytoplasmic ring of the basal body of the flagellum apparatus. Among all organisms, the most similar proteins outside the genus *Borrelia* are the FliL proteins of *T. pallidum*, *Treponema denticola*, and *Leptospira interrogans*, but the sequence identities with the proteins of these other spirochete species are less than 35%. In comparison, the FlaB protein of *B. burgdorferi* is 40% identical to the homologous flagellin proteins of *Treponema* spp. As a component of an immunoassay, FliL may show less antigenic cross-reactivity with the homologous proteins of other bacteria than has been the case with FlaB (59, 73).

Of all the newly identified Orfs, we paid most attention to BBK07 and BBK12. As determined by stringent criteria, these are predicted lipoproteins, and although the amino acid sequences are 88% identical, the ORFs are located several ORFs apart in the left arm of the lp36 plasmid. Comparison of the BBK07 and BBK12 gene sequences of strain B31 with the sequences of two other strains, 297 and N40, revealed >98% sequence identity between the strains for these sequences, an indication that a single example of each could be used to detect

antibodies to other strains of *B. burgdorferi*. Although BBK12 and, by inference, BBK07 are expressed by cells cultivated in the laboratory (Fig. 6), neither had previously been identified as an antigen. This may be attributable in part to the tendency of the lp36 plasmid to be lost sooner than other plasmids from *B. burgdorferi* during serial cultivation (7, 72, 79); thus, this plasmid may have frequently been absent from the lysates that investigators used for Western blotting and other fractionations in pursuit of diagnostic antigens. But another reason why BBK12 and BBK07 may have been overlooked is that these genes appear to be unique to *B. burgdorferi*. They have not been found to date in the two other major Lyme disease agents: *Borrelia afzelii* and *Borrelia garinii*. Using a DNA array comprising various lipoprotein genes of *B. burgdorferi*, Liang et al. did not find evidence of the BBK07 and BBK12 genes in either *B. afzelii* or *B. garinii* (55). Glockner et al. reported that they "did not find counterparts of the *B. burgdorferi* plasmids lp36 and lp38 or their respective gene repertoire in the *B. garinii* genome" (43). Our searches of all deposited GenBank sequences of *B. garinii* and *B. afzelii*, including two genomes of each species, likewise did not reveal a PF 59 ortholog. As determined by our analysis, the lp34 plasmid of *B. afzelii* has orthologs of *B. burgdorferi* ORFs in the order BBK01-BBK13-BBK15-BBK17-BBK21-BBK22-BBK23-BBK24, but BBK07 and BBK12 are absent from this and other replicons. If this genetic difference between LB species is confirmed, it suggests that BBK07 or BBK12 can be used in serological assays to distinguish *B. burgdorferi* infections from *B. afzelii* and *B. garinii* infections. These genetic distinctions between lineages may also provide insight into differences in pathogenesis and clinical manifestations between LB species.

APPENDIX

Estimation of the number of immunogenic Orfs. We assessed the size of the set of proteins that were immunogenic in *B. burgdorferi* infections by examining the relative amounts of binding for antibodies in panel 1 serum specimens and each of the 1,292 strain B31 Orfs. To do this, we determined the sign for each possible pair of sera in the panel. The member of a pair that had the higher intensity value for a given Orf was assigned a value of "1," and the pair member with lower reactivity was assigned a value of "0." As a hypothetical example, if serum *a* had an intensity value of 3,246 for Orf *x* and serum *b* had a value of 1,711 for Orf *x*, then serum *a* was assigned a value of "1" and serum *b* was assigned a value of "0" for the pairwise comparison in the matrix.

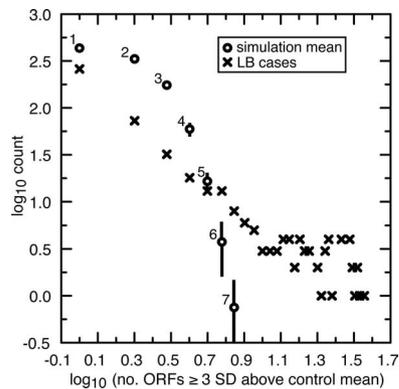


FIG. A1. Comparison of observed results with expected results from simulations: counts of Orfs in arrays with LB sera that were ≥ 3 SDs above the mean of control sera one or more times. Results for 39 panel 1 and 2 sera from patients with later LB are compared with mean counts (with 95% confidence intervals) for four simulation runs with random linkages. The numbers of Orfs that exceeded the 3-SD cutoff one to seven times are indicated next to the corresponding symbols for the random linkage simulation.

The complete pairwise results are shown in Table S3 in the supplemental material. Under the null hypothesis, a given serum sample would have the higher value of the pair in one-half of the comparisons, or 646 comparisons in this case. This was observed when controls were compared to controls, early-infection sera were compared to early-infection sera, and later-infection sera were compared to later-infection sera; the observed mean values were 646 (95% confidence interval, 482 to 810), 646 (95% confidence interval, 524 to 768), and 646 (95% confidence interval, 505 to 787), respectively. In contrast to these results for within-group pairs were the results for between-group pairs, e.g., a control serum and an LB serum. Table A1 summarizes the intergroup comparisons. Excess binding in the range from 100 to 200 Orfs was also observed with later-infection sera compared to early-infection sera. From these results, we estimated that the upper limit for immunogenic Orfs during human infection was ~ 200 , or $\sim 15\%$ of the 1,292 strain B31 Orfs on the array.

Simulation to establish a cutoff frequency. The mean and SD for each Orf with all control serum samples in each panel were determined. Then, for each Orf and for every serum sample in each panel, the number of SDs above or below the mean for the control sera in the same serum panel for the Orf in question was determined for normalization. For each sample all the Orfs that had array spots with values that were ≥ 3 SDs above the mean for the negative controls in the experiment for the given Orf were tabulated and summed. The frequencies of each Orf that appeared in this cumulative list were then determined. To provide an exact test of the significance of the counts that were obtained, we randomized the linkages of a given normalized value and an Orf and then likewise counted the times that an Orf was associated by chance with an SD that was ≥ 3 above the controls. This gave an estimate of the distribution under random conditions. Four replicates were performed, and the means were used to provide a distribution under the null hypothesis of random association between SD and Orf. Figure A1 shows the means and confidence intervals for the four replicates, i.e., what was expected under random

expectations. This is compared to what we observed with 39 sera from patients with later LB. (Frequency histograms of these distributions are shown in Fig. S1 in the supplemental material.)

ACKNOWLEDGMENTS

We thank Barbara Johnson and Martin Schriefer of the Centers for Disease Control and Prevention for providing panel 1 sera, Jozelyn Pablo and Chad Burk for cloning the *B. burgdorferi* vector set, and Rie Sasaki for preparing the purified recombinant proteins.

This research was supported by an institutional grant from the University of California Irvine (P.L.F. and A.G.B.); by NIH grants AI065359 (A.G.B.), AI072872 (P.L.F. and A.G.B.), LM007743 (M.A.K. and P.B.), and AR20358 (A.C.S.); by NSF grant MRI EIA-0321390 (M.A.K. and P.B.); by grants from the English, Bontner, Mitchell Foundation and the Eshe Fund (A.C.S.); and by patent royalties donated for research (A.G.B.).

REFERENCES

1. Aguerro-Rosenfeld, M. E., G. Wang, I. Schwartz, and G. P. Wormser. 2005. Diagnosis of Lyme borreliosis. *Clin. Microbiol. Rev.* **18**:484–509.
2. Altman, D. G., D. Machin, T. N. Bryant, and M. J. Gardner. 2000. *Statistics with confidence*. BMJ Books, London, United Kingdom.
3. Baldi, P., and H. G. Hatfield. 2002. *DNA microarrays and gene expression: from experiments to data analysis and modeling*. Cambridge University Press, Cambridge, United Kingdom.
4. Baldi, P., and A. D. Long. 2001. A Bayesian framework for the analysis of microarray expression data: regularized t-test and statistical inferences of gene changes. *Bioinformatics* **17**:509–519.
5. Barbour, A. G. 1984. Immunochemical analysis of Lyme disease spirochetes. *Yale J. Biol. Med.* **57**:581–586.
6. Barbour, A. G. 1984. Isolation and cultivation of Lyme disease spirochetes. *Yale J. Biol. Med.* **57**:521–525.
7. Barbour, A. G. 1988. Plasmid analysis of *Borrelia burgdorferi*, the Lyme disease agent. *J. Clin. Microbiol.* **26**:475–478.
8. Barbour, A. G., W. Burgdorfer, E. Grunwaldt, and A. C. Steere. 1983. Antibodies of patients with Lyme disease to components of the Ixodes dammini spirochete. *J. Clin. Investig.* **72**:504–515.
9. Barbour, A. G., S. F. Hayes, R. A. Heiland, M. E. Schrupf, and S. L. Tessler. 1986. A *Borrelia*-specific monoclonal antibody binds to a flagellar epitope. *Infect. Immun.* **52**:549–554.
10. Barbour, A. G., S. L. Tessler, and W. J. Todd. 1983. Lyme disease spirochetes and ixodid tick spirochetes share a common surface antigenic determinant defined by a monoclonal antibody. *Infect. Immun.* **41**:795–804.
11. Brinkman, M. B., M. McKevitt, M. McLoughlin, C. Perez, J. Howell, G. M. Weinstock, S. J. Norris, and T. Palzkill. 2006. Reactivity of antibodies from syphilis patients to a protein array representing the *Treponema pallidum* proteome. *J. Clin. Microbiol.* **44**:888–891.
12. Brisson, D., and D. E. Dykhuizen. 2004. ospC diversity in *Borrelia burgdorferi*: different hosts are different niches. *Genetics* **168**:713–722.
13. Brooks, C. S., P. S. Hefty, S. E. Jolliff, and D. R. Akins. 2003. Global analysis of *Borrelia burgdorferi* genes regulated by mammalian host-specific signals. *Infect. Immun.* **71**:3371–3383.
14. Bunikis, J., and A. G. Barbour. 2002. Laboratory testing for suspected Lyme disease. *Med. Clin. N. Am.* **86**:311–340.
15. Bunikis, J., U. Garpmo, J. Tsao, J. Berglund, D. Fish, and A. G. Barbour. 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.
16. Bunikis, J., C. J. Luke, E. Bunikiene, S. Bergström, and A. G. Barbour. 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.
17. Bunikis, J., L. Noppa, Y. Ostberg, A. G. Barbour, and S. Bergström. 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.
18. Bunikis, J., J. Tsao, C. J. Luke, M. G. Luna, D. Fish, and A. G. Barbour. 2004. *Borrelia burgdorferi* infection in a natural population of Peromyscus leucopus mice: a longitudinal study in an area where Lyme borreliosis is highly endemic. *J. Infect. Dis.* **189**:1515–1523.
19. Burgdorfer, W., A. G. Barbour, S. F. Hayes, J. L. Benach, E. Grunwaldt, and J. P. Davis. 1982. Lyme disease—a tick-borne spirochetosis? *Science* **216**:1317–1319.
20. Cadavid, D., P. M. Pennington, T. A. Kerentseva, S. Bergstrom, and A. G. Barbour. 1997. Immunologic and genetic analyses of VmpA of a neurotropic strain of *Borrelia turicatae*. *Infect. Immun.* **65**:3352–3360.
21. Casjens, S., N. Palmer, R. van Vugt, W. M. Huang, B. Stevenson, P. Rosa,

- R. Lathigra, G. Sutton, J. Peterson, R. J. Dodson, D. Haft, E. Hickey, M. Gwinn, O. White, and C. M. Fraser. 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.
22. Coleman, S. A., E. R. Fischer, D. C. Cockrell, D. E. Voth, D. Howe, D. J. Mead, J. E. Samuel, and R. A. Heinzen. 2007. Proteome and antigen profiling of *Coxiella burnetii* developmental forms. *Infect. Immun.* 75:290–298.
 23. Connolly, J. P., D. Comerci, T. G. Alefantis, A. Walz, M. Quan, R. Chafin, P. Grewal, C. V. Mujer, R. A. Ugalde, and V. G. DelVecchio. 2006. Proteomic analysis of *Brucella abortus* cell envelope and identification of immunogenic candidate proteins for vaccine development. *Proteomics* 6:3767–3780.
 24. Craft, J. E., D. K. Fischer, G. T. Shimamoto, and A. C. Steere. 1986. Antigens of *Borrelia burgdorferi* recognized during Lyme disease. Appearance of a new immunoglobulin M response and expansion of the immunoglobulin G response late in the illness. *J. Clin. Invest.* 78:934–939.
 25. Crother, T. R., C. I. Champion, J. P. Whitelegge, R. Aguilera, X. Y. Wu, D. R. Blanco, N. L. Miller, and M. A. Lovett. 2004. Temporal analysis of the antigenic composition of *Borrelia burgdorferi* during infection in rabbit skin. *Infect. Immun.* 72:5063–5072.
 26. Crotty, S., P. Felgner, H. Davies, J. Glidewell, L. Villarreal, and R. Ahmed. 2003. Cutting edge: long-term B cell memory in humans after smallpox vaccination. *J. Immunol.* 171:4969–4973.
 27. Daily, J. P., K. G. Le Roch, O. Sarr, D. Ndiaye, A. Lukens, Y. Zhou, O. Ndir, S. Mboup, A. Sultan, E. A. Winzeler, and D. F. Wirth. 2005. In vivo transcriptome of *Plasmodium falciparum* reveals overexpression of transcripts that encode surface proteins. *J. Infect. Dis.* 191:1196–1203.
 28. Das, S., D. Shraga, C. Gannon, T. T. Lam, S. Feng, L. R. Brunet, S. R. Telford, S. W. Barthold, R. A. Flavell, and E. Fikrig. 1996. Characterization of a 30-kDa *Borrelia burgdorferi* substrate-binding protein homologue. *Res. Microbiol.* 147:739–751.
 29. Davies, D. H., X. Liang, J. E. Hernandez, A. Randall, S. Hirst, Y. Mu, K. M. Romero, T. T. Nguyen, M. Kalantari-Dehaghi, S. Crotty, P. Baldi, L. P. Villarreal, and P. L. Felgner. 2005. Profiling the humoral immune response to infection by using proteome microarrays: high-throughput vaccine and diagnostic antigen discovery. *Proc. Natl. Acad. Sci. USA* 102:547–552.
 30. Davies, D. H., M. M. McCausland, C. Valdez, D. Huynh, J. E. Hernandez, Y. Mu, S. Hirst, L. Villarreal, P. L. Felgner, and S. Crotty. 2005. Vaccinia virus H3L envelope protein is a major target of neutralizing antibodies in humans and elicits protection against lethal challenge in mice. *J. Virol.* 79:11724–11733.
 31. Davies, D. H., D. M. Molina, J. Wrammert, J. Miller, S. Hirst, Y. Mu, J. Pablo, B. Unal, R. Nakajima-Sasaki, X. Liang, S. Crotty, K. L. Karem, I. K. Damon, R. Ahmed, L. Villarreal, and P. L. Felgner. 2007. Proteome-wide analysis of the serological response to vaccinia and smallpox. *Proteomics* 7:1678–1686.
 32. de Silva, A. M., and E. Fikrig. 1997. Arthropod- and host-specific gene expression by *Borrelia burgdorferi*. *J. Clin. Invest.* 99:377–379.
 33. Dressler, F., J. A. Whalen, B. N. Reinhardt, and A. C. Steere. 1993. Western blotting in the serodiagnosis of Lyme disease. *J. Infect. Dis.* 167:392–400.
 34. Engstrom, S. M., E. Shoop, and R. C. Johnson. 1995. Immunoblot interpretation criteria for serodiagnosis of early Lyme disease. *J. Clin. Microbiol.* 33:419–427.
 35. Eyles, J. E., B. Unal, M. G. Hartley, S. L. Newstead, H. Flick-Smith, J. L. Prior, P. C. Oyston, A. Randall, Y. Mu, S. Hirst, D. M. Molina, D. H. Davies, T. Milne, K. F. Griffin, P. Baldi, R. W. Titball, and P. L. Felgner. 2007. Immunodominant Francisella tularensis antigens identified using proteome microarray. Crown copyright 2007 Dstl. *Proteomics* 7:2172–2183.
 36. Feng, S., S. Das, T. Lam, R. A. Flavell, and E. Fikrig. 1995. A 55-kilodalton antigen encoded by a gene on a *Borrelia burgdorferi* 49-kilobase plasmid is recognized by antibodies in sera from patients with Lyme disease. *Infect. Immun.* 63:3459–3466.
 37. Feng, S., E. Hodzic, B. Stevenson, and S. W. Barthold. 1998. Humoral immunity to *Borrelia burgdorferi* N40 decorin binding proteins during infection of laboratory mice. *Infect. Immun.* 66:2827–2835.
 38. Forgher, M., R. Basu, K. Roychoudhury, S. Theinert, S. Roy, S. Sundar, and P. Walden. 2006. Mapping the antigenicity of the parasites in *Leishmania donovani* infection by proteome serology. *PLoS ONE* 1:e40.
 39. Fraser, C. M., S. Casjens, W. M. Huang, G. G. Sutton, R. Clayton, R. Lathigra, O. White, K. A. Ketchum, R. Dodson, E. K. Hickey, M. Gwinn, B. Dougherty, J. F. Tomb, R. D. Fleischmann, D. Richardson, J. Peterson, A. R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. van Vugt, N. Palmer, M. D. Adams, J. Gocayne, J. Weidman, T. Utterback, L. Watthey, L. McDonald, P. Artiach, C. Bowman, S. Garland, C. Fuji, M. D. Cotton, K. Horst, K. Roberts, B. Hatch, H. O. Smith, and J. C. Venter. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390:580–586.
 40. Gardy, J. L., M. R. Laird, F. Chen, S. Rey, C. J. Walsh, M. Ester, and F. S. Brinkman. 2005. PSORTb v. 2.0: expanded prediction of bacterial protein subcellular localization and insights gained from comparative proteome analysis. *Bioinformatics* 21:617–623.
 41. Gilmore, R. D., Jr., and M. L. Mbow. 1998. A monoclonal antibody generated by antigen inoculation via tick bite is reactive to the *Borrelia burgdorferi* Rev protein, a member of the 2.9 gene family locus. *Infect. Immun.* 66:980–986.
 42. Gilmore, R. D., Jr., R. L. Murphree, A. M. James, S. A. Sullivan, and B. J. Johnson. 1999. The *Borrelia burgdorferi* 37-kilodalton immunoblot band (P37) used in serodiagnosis of early Lyme disease is the *flaA* gene product. *J. Clin. Microbiol.* 37:548–552.
 43. Glockner, G., R. Lehmann, A. Romualdi, S. Pradella, U. Schulte-Spechtel, M. Schilhabel, B. Wilske, J. Suhnel, and M. Platzer. 2004. Comparative analysis of the *Borrelia garinii* genome. *Nucleic Acids Res.* 32:6038–6046.
 44. Haas, G., G. Karaali, K. Ebermayer, W. G. Metzger, S. Lamer, U. Zimny-Arndt, S. Diescher, U. B. Goebel, K. Vogt, A. B. Roznowski, B. J. Wiedenmann, T. F. Meyer, T. Aebischer, and P. R. Jungblut. 2002. Immunoproteomics of *Helicobacter pylori* infection and relation to gastric disease. *Proteomics* 2:313–324.
 45. Hansen, K., P. Hindersson, and N. S. Pedersen. 1988. Measurement of antibodies to the *Borrelia burgdorferi* flagellum improves serodiagnosis in Lyme disease. *J. Clin. Microbiol.* 26:338–346.
 46. Heikkila, T., I. Seppala, H. Saxen, J. Panelius, H. Yrjanainen, and P. Lahdenne. 2002. Species-specific serodiagnosis of Lyme arthritis and neuroborreliosis due to *Borrelia burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii* by using decorin binding protein A. *J. Clin. Microbiol.* 40:453–460.
 47. Hochberg, Y., and Y. Benjamini. 1990. More powerful procedures for multiple significance testing. *Stat. Med.* 9:811–818.
 48. Howe, T. R., L. W. Mayer, and A. G. Barbour. 1985. A single recombinant plasmid expressing two major outer surface proteins of the Lyme disease spirochete. *Science* 227:645–646.
 49. Jewett, M. W., K. Lawrence, A. C. Bestor, K. Tilly, D. Grimm, P. Shaw, M. VanRaden, F. Gherardini, and P. A. Rosa. 2007. The critical role of the linear plasmid lp36 in the infectious cycle of *Borrelia burgdorferi*. *Mol. Microbiol.* 64:1358–1374.
 50. Johnson, B. J., K. E. Robbins, R. E. Bailey, B. L. Cao, S. L. Sviat, R. B. Craven, L. W. Mayer, and D. T. Dennis. 1996. Serodiagnosis of Lyme disease: accuracy of a two-step approach using a flagella-based ELISA and immunoblotting. *J. Infect. Dis.* 174:346–353.
 51. Jwang, B., P. Dewing, E. Fikrig, and R. A. Flavell. 1995. The hook protein of *Borrelia burgdorferi*, encoded by the *flgE* gene, is serologically recognized in Lyme disease. *Clin. Diagn. Lab Immunol.* 2:609–615.
 52. Kowalczywska, M., F. Fenollar, D. Lafitte, and D. Raoult. 2006. Identification of candidate antigen in Whipple's disease using a serological proteomic approach. *Proteomics* 6:3294–3305.
 53. Lam, T. T., T. P. Nguyen, E. Fikrig, and R. A. Flavell. 1994. A chromosomal *Borrelia burgdorferi* gene encodes a 22-kilodalton lipoprotein, P22, that is serologically recognized in Lyme disease. *J. Clin. Microbiol.* 32:876–883.
 54. Lawrenz, M. B., J. M. Hardham, R. T. Owens, J. Nowakowski, A. C. Steere, G. P. Wormser, and S. J. Norris. 1999. Human antibody responses to VlsE antigenic variation protein of *Borrelia burgdorferi*. *J. Clin. Microbiol.* 37:3997–4004.
 55. Liang, F. T., F. K. Nelson, and E. Fikrig. 2002. DNA microarray assessment of putative *Borrelia burgdorferi* lipoprotein genes. *Infect. Immun.* 70:3300–3303.
 56. Liang, F. T., A. C. Steere, A. R. Marques, B. J. Johnson, J. N. Miller, and M. T. Philipp. 1999. Sensitive and specific serodiagnosis of Lyme disease by enzyme-linked immunosorbent assay with a peptide based on an immunodominant conserved region of *Borrelia burgdorferi* *vlsE*. *J. Clin. Microbiol.* 37:3990–3996.
 57. Lu, Z., M. I. Roche, J. H. Hui, B. Unal, P. L. Felgner, S. Gulati, G. Madico, and J. Sharon. 2007. Generation and characterization of hybridoma antibodies for immunotherapy of tularemia. *Immunol. Lett.* 112:92–103.
 58. Luft, B. J., P. D. Gorevic, W. Jiang, P. Munoz, and R. J. Dattwyler. 1991. Immunologic and structural characterization of the dominant 66- to 73-kDa antigens of *Borrelia burgdorferi*. *J. Immunol.* 146:2776–2782.
 59. Magnarelli, L. A., E. Fikrig, S. J. Padula, J. F. Anderson, and R. A. Flavell. 1996. Use of recombinant antigens of *Borrelia burgdorferi* in serologic tests for diagnosis of Lyme borreliosis. *J. Clin. Microbiol.* 34:237–240.
 60. McAtee, C. P., K. E. Fry, and D. E. Berg. 1998. Identification of potential diagnostic and vaccine candidates of *Helicobacter pylori* by "proteome" technologies. *Helicobacter* 3:163–169.
 61. McKevitt, M., M. B. Brinkman, M. McLoughlin, C. Perez, J. K. Howell, G. M. Weinstein, S. J. Norris, and T. Palzkill. 2005. Genome scale identification of *Treponema pallidum* antigens. *Infect. Immun.* 73:4445–4450.
 62. Meier, J. T., M. I. Simon, and A. G. Barbour. 1985. Antigenic variation is associated with DNA rearrangements in a relapsing fever *Borrelia*. *Cell* 41:403–409.
 63. Miller, J. C., and B. Stevenson. 2003. Immunological and genetic characterization of *Borrelia burgdorferi* BapA and EppA proteins. *Microbiology* 149:1113–1125.
 64. Nigrovic, L. E., and K. M. Thompson. 2007. The Lyme vaccine: a cautionary tale. *Epidemiol. Infect.* 135:1–8.
 65. Nowalk, A. J., R. D. Gilmore, Jr., and J. A. Carroll. 2006. Serologic proteome analysis of *Borrelia burgdorferi* membrane-associated proteins. *Infect. Immun.* 74:3864–3873.

66. Nowalk, A. J., C. Nolder, D. R. Clifton, and J. A. Carroll. 2006. Comparative proteome analysis of subcellular fractions from *Borrelia burgdorferi* by NEPHGE and IPG. *Proteomics* **6**:2121–2134.
67. Ojaimi, C., C. Brooks, D. Akins, S. Casjens, P. Rosa, A. Elias, A. Barbour, A. Jasinskas, J. Benach, L. Katonah, J. Radolf, M. Caimano, J. Skare, K. Swingle, S. Sims, and I. Schwartz. 2002. *Borrelia burgdorferi* gene expression profiling with membrane-based arrays. *Methods Enzymol.* **358**:165–177.
68. Padula, S. J., A. Sampieri, F. Dias, A. Szczepanski, and R. W. Ryan. 1993. Molecular characterization and expression of p23 (OspC) from a North American strain of *Borrelia burgdorferi*. *Infect. Immun.* **61**:5097–5105.
69. Panelius, J., P. Lahdenne, H. Saxen, T. Heikkila, and I. Seppala. 2001. Recombinant flagellin A proteins from *Borrelia burgdorferi* sensu stricto, *B. afzelii*, and *B. garinii* in serodiagnosis of Lyme borreliosis. *J. Clin. Microbiol.* **39**:4013–4019.
70. Porcella, S. F., C. A. Fitzpatrick, and J. L. Bono. 2000. Expression and immunological analysis of the plasmid-borne *mlp* genes of *Borrelia burgdorferi* strain B31. *Infect. Immun.* **68**:4992–5001.
71. Probert, W. S., and B. J. Johnson. 1998. Identification of a 47 kDa fibronectin-binding protein expressed by *Borrelia burgdorferi* isolate B31. *Mol. Microbiol.* **30**:1003–1015.
72. Purser, J. E., and S. J. Norris. 2000. Correlation between plasmid content and infectivity in *Borrelia burgdorferi*. *Proc. Natl. Acad. Sci. USA* **97**:13865–13870.
73. Rasiah, C., E. Schiltz, J. Reichert, and A. Vogt. 1992. Purification and characterization of a tryptic peptide of *Borrelia burgdorferi* flagellin, which reduces cross-reactivity in immunoblots and ELISA. *J. Gen. Microbiol.* **138**:147–154.
74. Revel, A. T., A. M. Talaat, and M. V. Norgard. 2002. DNA microarray analysis of differential gene expression in *Borrelia burgdorferi*, the Lyme disease spirochete. *Proc. Natl. Acad. Sci. USA* **99**:1562–1567.
75. Roberts, D. M., M. Caimano, J. McDowell, M. Theisen, A. Holm, E. Orff, D. Nelson, S. Wikel, J. Radolf, and R. T. Marconi. 2002. Environmental regulation and differential production of members of the Bdr protein family of *Borrelia burgdorferi*. *Infect. Immun.* **70**:7033–7041.
76. Roberts, W. C., B. A. Mullikin, R. Lathigra, and M. S. Hanson. 1998. Molecular analysis of sequence heterogeneity among genes encoding decorin binding proteins A and B of *Borrelia burgdorferi* sensu lato. *Infect. Immun.* **66**:5275–5285.
77. Sadziene, A., D. D. Thomas, V. G. Bundoc, S. C. Holt, and A. G. Barbour. 1991. A flagella-less mutant of *Borrelia burgdorferi*. Structural, molecular, and in vitro functional characterization. *J. Clin. Investig.* **88**:82–92.
78. Saeed, A., V. Sharov, J. White, J. Li, W. Liang, N. Bhagabati, J. Braisted, M. Klapa, T. Currier, M. Thiagarajan, A. Sturn, M. Snuffin, A. Rezansev, D. Popov, A. Ryltsov, E. Kostukovich, I. Borisovsky, Z. Liu, A. Vinsavich, V. Trush, and J. Quackenbush. 2003. TM4: a free, open-source system for microarray data management and analysis. *BioTechniques* **34**:374–378.
79. Schwan, T. G., W. Burgdorfer, and C. F. Garon. 1988. Changes in infectivity and plasmid profile of the Lyme disease spirochete, *Borrelia burgdorferi*, as a result of in vitro cultivation. *Infect. Immun.* **56**:1831–1836.
80. Schwan, T. G., and J. Piesman. 2000. Temporal changes in outer surface proteins A and C of the Lyme disease-associated spirochete, *Borrelia burgdorferi*, during the chain of infection in ticks and mice. *J. Clin. Microbiol.* **38**:382–388.
81. Schwan, T. G., J. Piesman, W. T. Golde, M. C. Dolan, and P. A. Rosa. 1995. Induction of an outer surface protein on *Borrelia burgdorferi* during tick feeding. *Proc. Natl. Acad. Sci. USA* **92**:2909–2913.
82. Steere, A. C., and S. M. Angelis. 2006. Therapy for Lyme arthritis: strategies for the treatment of antibiotic-refractory arthritis. *Arthritis Rheum.* **54**:3079–3086.
83. Steere, A. C., J. Coburn, and L. Glickstein. 2004. The emergence of Lyme disease. *J. Clin. Investig.* **113**:1093–1101.
84. Steere, A. C., G. McHugh, C. Suarez, J. Hoitt, N. Damle, and V. K. Sikand. 2003. Prospective study of coinfection in patients with erythema migrans. *Clin. Infect. Dis.* **36**:1078–1081.
85. Stevenson, B., J. L. Bono, T. G. Schwan, and P. Rosa. 1998. *Borrelia burgdorferi* Erp proteins are immunogenic in mammals infected by tick bite, and their synthesis is inducible in cultured bacteria. *Infect. Immun.* **66**:2648–2654.
86. Sundaresh, S., D. L. Doolan, S. Hirst, Y. Mu, B. Unal, D. H. Davies, P. L. Felgner, and P. Baldi. 2006. Identification of humoral immune responses in protein microarrays using DNA microarray data analysis techniques. *Bioinformatics* **22**:1760–1766.
87. Sundaresh, S., A. Randall, B. Unal, J. M. Petersen, J. T. Belisle, M. G. Hartley, M. Duffield, R. W. Titball, D. H. Davies, P. L. Felgner, and P. Baldi. 2007. From protein microarrays to diagnostic antigen discovery: a study of the pathogen *Francisella tularensis*. *Bioinformatics* **23**:i508–i518.
88. Tokarz, R., J. M. Anderton, L. I. Katona, and J. L. Benach. 2004. Combined effects of blood and temperature shift on *Borrelia burgdorferi* gene expression as determined by whole-genome DNA array. *Infect. Immun.* **72**:5419–5432.
89. Truchon, J.-F., and C. I. Bayly. 2007. Evaluating virtual screening methods: good and bad metrics for the “early recognition” problem. *J. Chem. Inf. Model.* **47**:488–508.
90. Tsao, J., A. G. Barbour, C. J. Luke, E. Fikrig, and D. Fish. 2001. OspA immunization decreases transmission of *Borrelia burgdorferi* spirochetes from infected *Peromyscus leucopus* mice to larval *Ixodes scapularis* ticks. *Vector Borne Zoonotic Dis.* **1**:65–74.
91. Ulvestad, E., A. Kanestrom, L. J. Sonstebj, R. Jureen, T. Omland, B. Edvardsen, J. Lundervik, E. Kristoffersen, and A. P. van Dam. 2001. Diagnostic and biological significance of anti-p41 IgM antibodies against *Borrelia burgdorferi*. *Scand. J. Immunol.* **53**:416–421.
92. Vapnik, V. 1995. The nature of statistical learning theory. Springer, New York, NY.
93. Vaz, A., L. Glickstein, J. A. Field, G. McHugh, V. K. Sikand, N. Damle, and A. C. Steere. 2001. Cellular and humoral immune responses to *Borrelia burgdorferi* antigens in patients with culture-positive early Lyme disease. *Infect. Immun.* **69**:7437–7444.
94. Wallich, R., M. M. Simon, H. Hofmann, S. E. Moter, U. E. Schaible, and M. D. Kramer. 1993. Molecular and immunological characterization of a novel polymorphic lipoprotein of *Borrelia burgdorferi*. *Infect. Immun.* **61**:4158–4166.
95. Wharton, M., T. L. Chorba, R. L. Vogt, D. L. Morse, and J. W. Buehler. 1990. Case definitions for public health surveillance. *MMWR Recommend. Rep.* **39**:1–43.
96. Wilske, B., V. Fingerle, P. Herzer, A. Hofmann, G. Lehnert, H. Peters, H. W. Pfister, V. Preac-Mursic, E. Soutschek, and K. Weber. 1993. Recombinant immunoblot in the serodiagnosis of Lyme borreliosis. Comparison with indirect immunofluorescence and enzyme-linked immunosorbent assay. *Med. Microbiol. Immunol.* **182**:255–270.
97. Zhang, H., A. Raji, M. Theisen, P. R. Hansen, and R. T. Marconi. 2005. *bdrF₂* of Lyme disease spirochetes is coexpressed with a series of cytoplasmic proteins and is produced specifically during early infection. *J. Bacteriol.* **187**:175–184.
98. Zückert, W. R., and J. Meyer. 1996. Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. *J. Bacteriol.* **178**:2287–2298.
99. Zückert, W. R., J. Meyer, and A. G. Barbour. 1999. Comparative analysis and immunological characterization of the *Borrelia* Bdr protein family. *Infect. Immun.* **67**:3257–3266.