

# Genetic Divergence and Evolutionary Instability in *ospE*-Related Members of the Upstream Homology Box Gene Family in *Borrelia burgdorferi* Sensu Lato Complex Isolates

SHIAN-YING SUNG, CRYSTAL P. LAVOIE, JASON A. CARLYON, AND RICHARD T. MARCONI\*

Department of Microbiology and Immunology, Medical College of Virginia of Virginia Commonwealth University, Richmond, Virginia 23298-0678

Received 27 April 1998/Returned for modification 26 June 1998/Accepted 10 July 1998

A series of related genes that are flanked at their 5' ends by a conserved upstream sequence element called the upstream homology box (UHB) have been identified in *Borrelia burgdorferi*. These genes have been referred to as the UHB or *erp* gene family. We previously demonstrated that among a limited number of *B. burgdorferi* isolates, the UHB gene family is variable in composition and organization. Prior to this report the UHB gene family in other species of the *B. burgdorferi* sensu lato complex had not been studied, and if this family is important in the pathogenesis or biology of the Lyme disease spirochetes, then a wide distribution among species and isolates of the *B. burgdorferi* sensu lato complex would be expected. To assess this, we screened for the UHB element by Southern hybridization and determined its restriction fragment length polymorphism (RFLP) patterns. The UHB element was found to be carried by all *B. burgdorferi* sensu lato complex species tested (*B. burgdorferi*, *B. garinii*, *B. afzelii*, *B. japonica*, *B. valaisiana* sp. nov., and *B. andersonii*), but the RFLP patterns varied widely at both the inter- and intraspecies levels. Variation in both the number and size of the hybridizing restriction fragments was evident. PCR analyses also revealed the presence of polymorphic, *ospE*-related alleles in many isolates. Sequence analyses identified the molecular basis of the polymorphisms as being primarily insertions and deletions. Sequence variation and the insertions and deletions were found to be clustered in two distinct domains (variable domains 1 and 2). In many isolates variable domain 1 is flanked by direct repeat elements, some as long as 38 bp. Computer analyses of the deduced amino acid sequences encoded within variable domain 1 predict them to be hydrophilic, surface exposed, and antigenic. The analyses conducted here suggest that the UHB gene family, as evidenced by the variable UHB RFLP patterns, is not evolutionarily stable and that the polymorphic *ospE* alleles are derived from a common ancestral gene which has been modified through mutation or recombination events. The characterization of *ospE*-related genes of the UHB gene family among *B. burgdorferi* sensu lato species will prove important in attempts to construct a model for UHB gene family organization and in deciphering the role of the UHB gene family in the biology and pathogenesis of the Lyme disease spirochetes.

Numerous studies have demonstrated that the *Borrelia burgdorferi* genome carries repeated sequences, numerous gene families, and multiple copies of closely related plasmids (4–6, 9, 11, 20, 22, 24, 25, 29). *B. burgdorferi* B31T may carry as many as seven related 32-kb circular plasmids (cp32s) (20) that carry some members of the upstream homology box (UHB) gene family (also referred to as the *erp* gene family) (6, 20, 25). The UHB gene family is defined by the presence of a conserved upstream sequence called the UHB element (1, 14, 20, 25–27). Nucleotide sequence identity values for the UHB elements from different isolates and from different UHB gene family members range from 81 to 100% (20). The conservation of this putative control element may suggest that the gene family is responsive to common regulatory factors or environmental regulatory signals and could constitute a regulon.

*ospE* and *ospF*, the first UHB gene family members to be identified, were found to exist as an operon in *B. burgdorferi* N40 (14). However, in other *B. burgdorferi* isolates these genes are not linked (1, 20). Some of the *ospE*-related genes that have been described exist in operons with downstream genes

distinct from *ospF* (25), and some isolates carry paralogs of *ospF* that exist independently of *ospE* (1, 20). The emerging picture is that the UHB gene family is highly variable and more complex than previously recognized. UHB gene family members that have been identified to date include *ospEF* (14), *ospEi* and *ospFi* (20), *pG* (27), *bbk2.10* (1), the *erp* genes (6, 25), and *p21* (26). These genes were identified in a variety of isolates; hence, it remains to be determined if all or just a subset of these genes are carried by individual isolates. Among the genes that are flanked at their 5' ends by UHB elements, there is a subgroup which exhibits high homology to *ospE*. This UHB gene family subgroup includes *ospEi*, *erpA*, *erpI*, *erpC*, and *p21*. The nucleotide identity values of these genes with *ospE* range from 85 to 100% (a summary of sequence identities and other properties of these genes is presented in Table 1). We refer to these genes collectively as *ospE*-related genes or *ospE* paralogs.

In this study we have characterized the restriction fragment length polymorphism (RFLP) patterns, distribution, and copy number of the UHB element and have assessed the genetic variability of *ospE*-related genes among isolates through PCR and DNA sequence analyses. Although UHB elements were detected in all *B. burgdorferi* sensu lato species tested, the composition of the gene family appears to vary among isolates. PCR, sequence, and Southern analyses of *ospE* paralogs have revealed that polymorphisms in these genes are localized in

\* Corresponding author. Mailing address: Department of Microbiology and Immunology, Sanger Hall, Medical College of Virginia of Virginia Commonwealth University, 1101 East Marshall St., Richmond, VA 23298-0678. Phone: (804) 828-3779. Fax: (804) 828-9946. E-mail: rmarconi@hsc.vcu.edu.



TABLE 1. Sequence accession numbers and features of UHB gene family members

Accession number and gene name (reference)	Isolate of origin <sup>a</sup>	Gene feature(s) <sup>b</sup>
L13924 <i>ospE</i> (14)	N40	First gene of the <i>ospEF</i> operon in <i>B. burgdorferi</i> isolate N40
L13925 <i>ospF</i> (14)	N40	Second gene of the <i>ospEF</i> operon in isolate N40
L78248 <i>ospEi</i> (20)	B31	91% nt identity with N40 <i>ospE</i> , not 3' flanked by <i>ospF</i>
L78250 <i>ospEi</i> (20)	N40CH	87% nt identity with N40 <i>ospE</i> , not 3' flanked by a conserved <i>ospF</i>
U44912 <i>ospE</i> (25)	B31T	Flanked 3' by <i>erpB</i> in some isolates, 85% nt identity to <i>ospE</i>
U78764 <i>erpA2</i>	B31T	Identical in sequence to <i>erpA</i> but carried on a different csp32 in isolate B31
U44912 <i>erpB</i> (25)	B31T	Resides 3' of <i>erpA</i> in some isolates, 36% nt identity to <i>ospF</i>
U44914 <i>erpC</i> (25)	B31T	Flanked 3' by <i>erpD</i> in some isolates, 88% nt identity to <i>ospE</i>
U44914 <i>erpD</i> (25)	B31T	Resides 3' of <i>erpC</i> in some isolates, 38% nt identity to <i>ospF</i>
U72996 <i>erpI</i> (6)	B31T	Flanked 3' by <i>erpI</i> , 100% nt identity to <i>erpA</i>
U72997 <i>erpK</i> (6)	B31T	Peripheral number of the UHB gene family exhibiting 54 and 68% nt identity with <i>ospE</i> and <i>ospF</i> , respectively
U72996 <i>erpI</i> (6)	B31T	Resides 3' of <i>erpI</i> in isolate B31T, 100% nt identity to <i>erpB</i>
L78249 <i>ospE</i> (20)	297CH	86% nt identity with N40 <i>ospE</i> , not 3' flanked by a conserved <i>ospF</i>
X82409 <i>pC</i> (27)	Z57	Immediately upstream from the <i>bapA</i> ( <i>bapA</i> is homologous to <i>ospA</i> [7])
X82409 <i>bapA</i> (27)	Z57	Downstream from <i>pC</i> , 80% amino acid similarity with <i>ospA</i> , not 3' flanked by a UHB element
L32797 <i>p21</i> (26)	N40	86 and 84% nt identity to N40 <i>ospE</i> and B31 <i>ospEi</i> , respectively
U18292 <i>bbk2.10</i> (1)	Z97	77% nt identity to Z57 <i>pG</i> , 3' flanked by <i>bapA</i>
U19754 <i>ospF</i> (1)	Z97	92 and 98% nt identity to N40 <i>ospF</i> and N40CH <i>ospF</i> , respectively
L78245 <i>ospF</i> (20)	297CH	100% identity with N40CH <i>ospF</i> , 5' flanked by a UHB element, not preceded by <i>ospE</i>
L78251 <i>ospF</i> (20)	N40CH	91% nt identity with N40 <i>ospF</i> , 5' flanked by a UHB element, not preceded by <i>ospE</i>
L79960 <i>ospF</i> (20)	CA12	Full-length <i>ospEF</i> operon, <i>ospE</i> and <i>ospF</i> exhibit 79 and 84% nt identity to N40 <i>ospE</i> and <i>ospF</i> , respectively
U30617 <i>bbk2.11</i> (1)	297	<i>ospF</i> variant

<sup>a</sup> All isolates are *B. burgdorferi*.<sup>b</sup> All genes or operons, unless otherwise indicated, are flanked at their 5' ends by a highly conserved UHB element.

distinct domains which computer analyses predict to encode hydrophilic, surface-exposed, and antigenic amino acid sequences. Genetic rearrangement and recombination appear to have contributed to the variable organization of these evolutionarily unstable genes. It is our hypothesis that recombination in and among *ospE*-related genes could result in a continually evolving, antigenically variable group of surface-exposed proteins. It is conceivable that these proteins could influence the host-pathogen interaction.

#### MATERIALS AND METHODS

**Bacterial cultivation and DNA isolation.** Bacterial isolates (Table 2) were cultivated in BSK-H medium (Sigma) supplemented with 0.6% (v/vol) bovine serum (Sigma) at 37°C, harvested by centrifugation, and washed with phosphate-buffered saline (pH 7.0). Two different *B. burgdorferi* B31 cultures were analyzed. *B. burgdorferi* B31T (a cloned isolate) was provided by Sherwood Casjens (University of Utah), and a second population, designated B31, was provided by Tom Schwan (Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health). DNA was isolated from all isolates listed in Table 2 as previously described (18, 21).

**RFLP pattern determination and Southern blot hybridizations.** *Hae*III-digested DNA was fractionated in 0.8% (w/vol) GTG-agarose gels (United States Biochemical), vacuum blotted onto Hybond N membranes (Amersham), UV cross-linked by using a GS-GeneLinker (Bio-Rad), and hybridized with the 5'-end-labeled oligonucleotides described in Table 3. Oligonucleotides were labeled at their 5' OH groups by using polynucleotide kinase and [ $\gamma$ -<sup>32</sup>P]ATP (6,000 Ci/mmol<sup>2</sup>; NEN-DuPont). Hybridizations with oligonucleotide probes were conducted at temperatures of 32 to 42°C in a Hybrid hybridization oven (Labnet). The hybridization buffer consisted of 0.2% (w/vol) bovine serum albumin, 0.2% (w/vol) polyvinyl-pyrrolidone (molecular weight, 40,000), 0.2% (w/vol) Ficoll (molecular weight, 400,000), 50 mM Tris-HCl (pH 7.5), 0.1% (w/vol) sodium pyrophosphate, 1% (w/vol) sodium dodecyl sulfate (SDS), 10% (v/vol) dextran sulfate, 100 µg of bovine serum albumin (BSA) ml<sup>-1</sup>, and 1 M NaCl. Two 10-min washes with 2× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) at 0.1% SDS and a 1-h wash with 0.2× SSC-0.1% SDS were performed at temperatures ranging from 32 to 42°C, depending on the probe. A final 5-min wash was done at room temperature with 0.2× SSC-0.1% SDS with vigorous shaking.

**PCR.** PCR amplification was performed with *Taq* polymerase (Promega) in 30-µl reaction volumes as previously described (19). Briefly, cycling conditions were 95°C for 3 min followed by 30 cycles of 95°C for 1 min, 50°C for 1 min, and 72°C for 1.5 min. Reaction mixtures were overlaid with light mineral oil

(Rite-Aid Pharmacy). Primer and oligonucleotide probe sequences and their intended target sites are listed in Table 3. PCR products were analyzed by electrophoresis in 1.2% GTG-agarose gels in TAE buffer (Tris-acetate [pH 8.5], 2 mM EDTA).

**Cloning and DNA sequence analysis.** Selected PCR products or recombinant clones were purified by using Wizard columns (Promega) and directly sequenced by using end-labeled primers and the *fast* DNA sequencing kit (Promega). Sequencing reactions were analyzed in 6% polyacrylamide-8 M urea gels at 85 W. The gels were transferred directly onto Whatman 3MM paper, wrapped in cellophane, and exposed to film for 1 to 3 h with intensifying screens. Sequences were analyzed by using the Wisconsin Sequence Analysis Package, version 9.0 (Genetics Computer Group, Madison, Wis.). To determine pairwise identity and similarity values, the GAP program was run with default parameters. Multiple sequence alignments were generated by using PILEUP and then manually refined. To conduct evolutionary analyses and construct phylogenies, the multiple sequence alignments were used as the input files for the DISTANCES program, which generates a matrix of evolutionary distances. Unrooted distance trees were determined. These distances were then used as input for the GROWTREE program to generate phylogenies. The neighbor-joining method was used, and negative branch lengths were not allowed.

Nucleotide sequence accession numbers. The PCR amplicons obtained with the *uhb*(+)E470(-) primer set from various isolates were assigned accession numbers AF029091 through AF029192. The sequence of the amplicon from *B. burgdorferi* 1D1 was submitted at a later date and was assigned the accession number AF059178.

#### RESULTS

**RFLP pattern analysis of the UHB element.** A previous analysis of six *B. burgdorferi* isolates demonstrated that UHB RFLP patterns are variable and suggested that the complement of UHB gene family members varies among *B. burgdorferi* isolates (20). To date, the presence and composition of the UHB gene family in other species of the *B. burgdorferi* sensu lato complex have not been assessed. To address this and to further analyze genetic diversity within *B. burgdorferi*, we screened for the UHB element in other *B. burgdorferi* sensu lato isolates by Southern hybridization and determined the UHB RFLP patterns by using *Hae*III-digested DNA. To verify

TABLE 2. Description of isolates utilized in this study and summary of PCR analyses

Species and isolate(s)	Origin <sup>a</sup>	In vitro passage <sup>b</sup>	PCR results with <i>uhb</i> (+)E470(-) primer set <sup>c</sup>
<i>B. burgdorferi</i>			
297, 297CH	Human CSF, CT	U (1), 7 (1)	+ (mp), +
N40 <sup>d</sup> , N40CH, B31, B31T <sup>d</sup>	<i>I. scapularis</i> , NY	L (1), 7 (1), L (1), L (1)	+, +, +, +
R100, T2, 25015	<i>I. scapularis</i> , NY	4 (1), 8, 6 (1)	+ (p), + (mp), +
K42, CA3, CA12	<i>I. hexagoni</i> , CA	11, 11, 25	+, +, +
CA4, CA7, CA8, CA9	<i>I. pacificus</i> , CA	5, 7, 8, 3	+, +, +, +
CA13	<i>I. scapularis</i> , CA	9	+
LP3, LP4, LP5, LP7	Human, CT	3, 3, 3, 3	+, +, +, +
272	Human EM, USA	28	+ (mp)
NY186	Human EM, NY	20	+
HBN <sup>c</sup>	Human blood, CA	2	+
JD1 <sup>d</sup>	<i>I. scapularis</i> , MA	4 (1)	+ (p)
VS134, VS307	<i>I. ricinus</i> , Switzerland	8, 7	+/-, +
<i>B. garinii</i>			
IP90, IP89	<i>I. persulcatus</i> , Russia	H, 4	+ (p), +
VS102, VSBP	<i>I. ricinus</i> , Switzerland	9, 8	+
G2	Human CSF, Germany	H	-
153	<i>I. ricinus</i> , France	11	+/-
FRG, Pbl, N34	<i>I. ricinus</i> , Germany	11, U, H	+, -, +, +/-
G25	<i>I. ricinus</i> , Sweden	H	+/-
BS-92, B4-91	Human EM, Norway	L, L	+
B4-87, B6-91	<i>I. ricinus</i> , Norway	9	+/-, +
<i>B. valaisiana</i> sp. nov. VS116	<i>I. ricinus</i> , Switzerland	9	+
<i>B. afzelii</i>			
J1	<i>I. persulcatus</i> , Japan	U	+
Pko	Human CSF, Germany	3	+
IP21, IP3	<i>I. persulcatus</i> , Russia	H, H	+ (p), +
Bo23, P6au <sup>d</sup>	Human EM, Germany	U, L	+, + (mp)
Pko <sup>d</sup>	Human EM, Switzerland	U	+
UMO1, ECM1, VS461	Human skin, Sweden	U, U, 9	+ (p), +, +/-
<i>B. andersonii</i>			
21038 <sup>d</sup>	<i>I. dentatus</i> , NY	L	+ (p)
19857	Rabbit kidney, NY	22	+
<i>B. japonica</i> IKA2, HO14	<i>I. ovatus</i> , Japan	U, U	+ (mp), -

<sup>a</sup> CSF, cerebrospinal fluid; EM, erythema migrans; CT, Connecticut; NY, New York; CA, California; USA, United States; MA, Massachusetts.<sup>b</sup> U, unknown passage history; L, passaged in vitro fewer than 15 times; H, passaged more than 30 times; I, confirmed to be infectious.<sup>c</sup> +, multiple bands were obtained; -, amplicons of a size different (polymorphic) from that predicted were obtained; +, strong amplification; +/-, weak amplification; -, no amplification.<sup>d</sup> The isolate was cloned by plating.

that the DNA had been completely digested, we performed Southern hybridization with a probe targeting the single-copy 16S rRNA gene (*rrs*) (23). All isolates yielded a single hybridizing band, indicating complete digestion (data not shown). To

screen for the UHB element, two different probes were used; the *uhb*(+) oligonucleotide and a UHB element-targeting, PCR-generated probe (472 bp). Oligonucleotide probe binding sites are presented in Table 3. The PCR probe was gener-

TABLE 3. Oligonucleotide probe and primer sequences and target sites<sup>a</sup>

Primer	Sequence (5'-3')	Intended target site positions
<i>uhb</i> (+)	GTGCGTTAAATTTACATTTGG	-128 to -107 of the <i>ospEF</i> UHB element <sup>b</sup> ( <i>B. burgdorferi</i> N40)
<i>uhb</i> (-)	TCCTCTCTCTTACAGGCTTAC	-426 to -406 of the <i>erpAB</i> UHB element ( <i>B. burgdorferi</i> B31)
E13(+)	ATGTTTATATTTGCTGTTT	13 to 40 of <i>ospE</i> ( <i>B. burgdorferi</i> N40)
E46(+)	CTTATATGCTGCTTCAAG	46 to 63 of <i>ospE</i> ( <i>B. burgdorferi</i> N40)
E46(-)	TTCTGACAGCACTTAAG	46 to 63 of <i>ospE</i> ( <i>B. burgdorferi</i> N40)
E70(-)	ACTTCATATGATGACGAAGT	70 to 90 of <i>ospE</i> ( <i>B. burgdorferi</i> N40)
E31(-)	AAACTGTTTAAATGATTC	310 to 327 of <i>ospE</i> ( <i>B. burgdorferi</i> N40)
C24(+)	CTAGTATATTCCTTCAAG	470 to 490 of <i>ospE</i> ( <i>B. burgdorferi</i> N40)
C24(-)	TGCGATATGATCAGAGTTCT	241 to 261 of <i>erpC</i> ( <i>B. burgdorferi</i> B31)
p21-105 (-)	CGCATCTTCTGCTTCACT	105 to 124 of p21
A468(-)	GTCCTCCCGCCAGCTTTA	450 to 468 of <i>ospE</i> ( <i>B. burgdorferi</i> B31)

<sup>a</sup> Primer designations are based on the numbering presented in the original descriptions of each individual gene (not on the alignments presented in this report). The data presented in this report demonstrate that several of the probes can hybridize with more than one *ospE* variant; hence, the probe designations themselves do not imply that the probes are absolutely variant specific. The designations (+) and (-) indicate that the primers are positive- or negative-strand primers, respectively. The sequences of the (-) primers are the inverse complement of the coding sequence.

<sup>b</sup> A conserved binding site for this primer also resides upstream of other UHB gene family members at an analogous position.

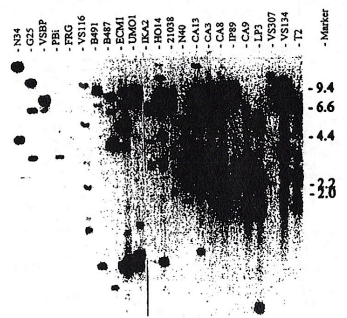


FIG. 1. RFLP patterns of the UHB element among *B. burgdorferi* sensu lato complex isolates. *Hae*III-digested DNA was transferred onto a Hybond N membrane and hybridized with the *uhb*(+) oligonucleotide under conditions described in the text. The isolates analyzed are indicated above the lanes, and molecular size standards (in kilobases) are indicated on the right. Isolates N24, G25, V58P, PBI, FRG, B491, and B487 are *B. garinii*; VS116 is *B. valaisiana*; UMO1 and UMO2 are *B. afzelii*; UMO3 and UMO4 are *B. japonica*; UMO5 is *B. andersonii*; and UMO6, UMO7, UMO8, UMO9, UMO10, UMO11, UMO12, UMO13, UMO14, UMO15, UMO16, UMO17, UMO18, UMO19, UMO20, UMO21, UMO22, UMO23, UMO24, UMO25, UMO26, UMO27, UMO28, UMO29, UMO30, UMO31, UMO32, UMO33, UMO34, UMO35, UMO36, UMO37, UMO38, UMO39, UMO40, UMO41, UMO42, UMO43, UMO44, UMO45, UMO46, UMO47, UMO48, UMO49, UMO50, UMO51, UMO52, UMO53, UMO54, UMO55, UMO56, UMO57, UMO58, UMO59, UMO60, UMO61, UMO62, UMO63, UMO64, UMO65, UMO66, UMO67, UMO68, UMO69, UMO70, UMO71, UMO72, UMO73, UMO74, UMO75, UMO76, UMO77, UMO78, UMO79, UMO80, UMO81, UMO82, UMO83, UMO84, UMO85, UMO86, UMO87, UMO88, UMO89, UMO90, UMO91, UMO92, UMO93, UMO94, UMO95, UMO96, UMO97, UMO98, UMO99, UMO100 are *B. burgdorferi*.

ated from *B. burgdorferi* B31T by using the *uhb*(+) and *E46*(-) primers. Both the *uhb*(+) oligonucleotide (Fig. 1) and the PCR-generated probe (data not shown) hybridized with DNAs from all *B. burgdorferi* sensu lato species (*B. burgdorferi*, *B. garinii*, *B. afzelii*, *B. andersonii*, *B. japonica*, and *B. valaisiana*). As with the oligonucleotide probe, no two isolates exhibited the same RFLP pattern when the PCR-generated probe was used. However, for each individual isolate these two probes yielded nearly identical RFLP patterns. When differences were observed, they were mainly in the form of differences in hybridization intensity of the probes with some bands. The two different UHB-targeting probes were used in these analyses because while oligonucleotide probes are best suited for copy number determination, it is possible that some UHB copies could go undetected as a result of sequence divergence within the oligonucleotide target site. Since the PCR probe carries an entire copy of the UHB element (i.e., that region extending from the 5' end of *ospE* upstream to the start codon of the oppositely oriented ORF6 [6]), minor sequence divergence would not prevent its hybridization. For *B. garinii* isolates, identical hybridization profiles were obtained with both probes. This demonstrates that the low UHB copy number in these species, inferred from hybridization with the *uhb*(+) oligonucleotide, is accurate. Hence, it can be concluded that the copy number of the UHB element varies among isolates. The number of UHB elements detected in *B. garinii* isolates was lowest, ranging from one to three. In contrast, *B. burgdorferi* and *B. afzelii* isolates carry four or more UHB elements. Since the RFLP patterns differ in both the number and size of the hybridizing restriction fragments, it can be concluded that the pattern differences are not due solely to loss of some plasmids, since this would only decrease the number of hybridizing bands. Since no discernible conservation of UHB RFLP patterns was observed among isolates of the same species, it can

be concluded that the organization of these genes is not reflective of phylogenetic relationships. Hence, UHB gene family composition and organization appear to have been influenced by recent molecular events such as mutation and/or recombination and plasmid loss or acquisition.

PCR analyses of *ospE*-related genes. As one step towards assessing possible genetic heterogeneity in *ospE*-related genes, PCR analyses were performed on 56 *B. burgdorferi* sensu lato isolates with the *uhb*(+)-*E470*(-) primer set (Table 2). PCR screening can reveal polymorphisms (insertions or deletions) that might not be evident by Southern blotting. The primers, *E470*(-) and *uhb*(+), have conserved binding sites in *ospE*, *erpA*, and *erpI*. Hence, if all three of these genes are present in the genome, each could be amplified by the primer set. It should be noted that a search of the database indicates that two separate but identical copies of *erpA* (designated *erpA* and *erpA2*; for accession numbers, see Table 1) are apparently carried by *B. burgdorferi* B31T. Since these alleles and *erpI* are identical in sequence, we did not attempt to differentiate between them. *uhb*(+)-*E470*(-) amplicons were obtained from 91% of the isolates tested (51 of 56), with some isolates yielding multiple bands (Table 2; representative data are also presented in Fig. 2). Many of the amplicons were polymorphic, being of a size not consistent with that predicted for *ospE*, *erpA*, *erpI*, or any other known *ospE* paralog, suggesting that there may be other polymorphic *ospE*-related genes carried by some *B. burgdorferi* sensu lato isolates. Hybridization and DNA sequence analyses designed to identify the molecular basis of these possible gene polymorphisms are described in detail below.

Southern hybridization analyses of *uhb*(+)-*E470*(-) PCR amplicons with probes targeting previously identified *ospE*-related genes. Since many of the *uhb*(+)-*E470*(-) amplicons were polymorphic in size, the identity of the amplified genes could not be inferred based upon migration in agarose gels. In addition, since *ospE* and the *ospE* paralogs *erpA* and *erpI* possess *uhb*(+) and *E470*(-) binding sites at analogous positions and would yield amplicons of the same size, it remained to be determined which of these genes were amplified. To aid in the identification of the polymorphic amplicons and to determine specifically which genes were being amplified, hybridization analyses of the amplicons were conducted. To determine if *erpA* (or *erpI*) was amplified, we used the *erpA* (*erpI*)-specific A468(-) oligonucleotide as a probe. The A468(-) probe hybridized with amplicons derived from only a few *B. burgdorferi* isolates (Fig. 2). The absence of hybridization of this probe with the amplicons implies that *erpA* (both copies) and *erpI* are possibly absent from the genome. Alternatively, it is conceivable, although unlikely, that all three *erpA*-related alleles (both copies of *erpA* and *erpI*) exhibit sequence divergence at the probe binding site in all A468(-) hybridization-negative isolates and therefore do not hybridize with the probe. The possible absence of these 5' UHB-flanked alleles from most isolates is indirectly supported by the lower UHB element copy number observed in most isolates compared to that in *B. burgdorferi* B31T (6, 9).

The blot probed with the A468(-) primer was stripped and sequentially probed with other *ospE* paralog-targeting probes. The E13(+), which targets the relatively conserved putative leader peptide of all *ospE*-related genes (except *erpC*, which is divergent within the probe binding site), hybridized with amplicons from 23 of the 28 PCR products tested. Some amplicons did not hybridize even under low-stringency conditions, indicating sequence divergence at their 5' termini. The E46(+), which targets within the 5' ends of *ospE*, *erpI*, and *erpA*, hybridized with amplicons derived from nine *B.*

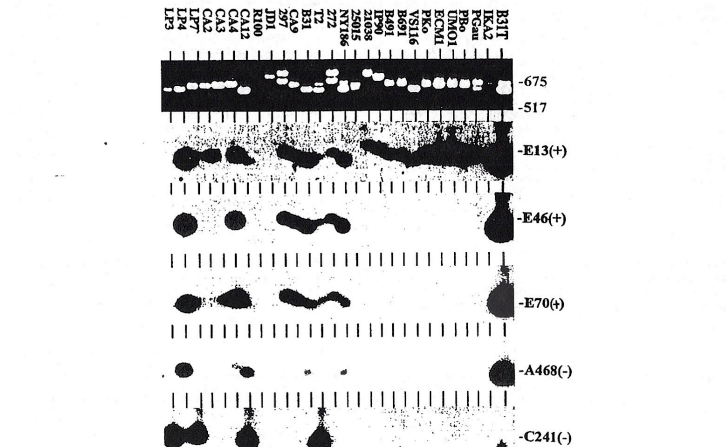


FIG. 2. PCR analyses of *ospE* and hybridization analyses of the amplicons with oligonucleotide probes targeting various *ospE* paralogs. PCR with the *uhb*(+)-*E470*(-) primer set was performed on various isolates. Ten microliters of each reaction mixture was analyzed in a 1.2% agarose gel and then stained with ethidium bromide. (Top panel) Representative PCR data. (Lower panels) Hybridization results with the amplicons from the top panel. The isolates analyzed are indicated at the top, and the oligonucleotide probes used are indicated at the right. Isolates L3, L4, L5, L6, L7, L8, L9, L10, L11, L12, L13, L14, L15, L16, L17, L18, L19, L20, L21, L22, L23, L24, L25, L26, L27, L28, L29, L30, L31, L32, L33, L34, L35, L36, L37, L38, L39, L40, L41, L42, L43, L44, L45, L46, L47, L48, L49, L50, L51, L52, L53, L54, L55, L56, L57, L58, L59, L60, L61, L62, L63, L64, L65, L66, L67, L68, L69, L70, L71, L72, L73, L74, L75, L76, L77, L78, L79, L80, L81, L82, L83, L84, L85, L86, L87, L88, L89, L90, L91, L92, L93, L94, L95, L96, L97, L98, L99, L100 are *B. burgdorferi*; 21038 is *B. andersonii*; IP90, B491, and B591 are *B. garinii*; VS116 is *B. valaisiana*; Pko, UMO1, EGM1, Pko, and Pgu are *B. afzelii*; and IK2 is *B. japonica*.

*burgdorferi* isolates but not with amplicons derived from other species, indicating that this region of the amplicons is variable among isolates, with divergence being pronounced across species lines. The C241(-) and p21-105(-) probes, which target *erpC* and *p21*, respectively, were not predicted to hybridize with the *uhb*(+)-*E470*(-) amplicons, since both the *erpC* and *p21* genes lack an *E470*(-) binding site and therefore should not be amplified. However, hybridization was observed with several *B. burgdorferi*-derived amplicons (*p21* data are not shown). It is possible either that some *erpC* or *p21* genes possess an *E470*(-) binding site or that sequences thought to be unique to *erpC* or *p21* are present in other *ospE* paralogs. These data raised the possibility that there may be *ospE* paralogs that are composite or mosaic genes.

Although the amplicons were screened by hybridization with probes targeting all of the identified *ospE* paralogs, some amplicons did not hybridize with this collection of probes. Examples of hybridization-negative amplicons include those derived from *B. burgdorferi* R100 and J1 and the larger of the two bands from *B. burgdorferi* 297 and 272 (all of which were identical in size). The absence of hybridization with the collection of probes used suggests that the amplified genes are divergent in sequence from other *ospE* paralogs. It can be concluded from these analyses that the *uhb*(+)-*E470*(-) amplicons are derived from a variable gene or group of genes and

that there are additional *ospE* paralogs carried by some isolates that have not yet been characterized and reported on in the literature. To assess the molecular variation in these amplicons, several were cloned or purified and their sequences were determined.

DNA sequence analyses of the *uhb*(+)-*E470*(-) amplicons. Selected *uhb*(+)-*E470*(-) amplicons were purified and sequenced. The determined sequences extend from approximately 120 bases upstream of the translational start codons to within approximately 40 bases 5' of the stop codon (as inferred from the location of the known stop codon of *ospE*). Thus, the determined sequences are partial and are missing a small segment of their 3' ends. To assess the relatedness between the coding regions of the amplicons, we conducted pairwise sequence comparisons by using the GAP program (default parameters). These nucleotide identity values ranged from 67.25 to 100% (Table 4). Excluding the amplicon sequence from *B. burgdorferi* J1, which was found to be the most peripheral of the nucleotide sequences, nucleotide identity values at the intraspecific level ranged from 83 to 100%. The average nucleotide identity value for these sequences compared with the corresponding coding sequences of *ospE*, *erpA*, *p21*, and *erpC* (from *B. burgdorferi* N40 or B31T) were 86.4, 87.6, 85.5, and 86.1%, respectively, indicating that the amplified genes are evolutionarily equidistant from each of the previously de-



TABLE 4. Nucleotide sequence identity and amino acid similarity values of the uhh(+)-E470(-) amplicons from *B. burgdorferi* sensu lato isolates and comparison with other UHB-flanked genes

Isolate or gene	% Nucleotide identity or amino acid similarity <sup>a</sup> :														
	21038	781	CA2	CA3	CA4	CA5	epv1	epv2	IP90	LP4	LP8	ospE	p21	V5116	IKAc23
21038	78.42	76.86	76.02	77.87	77.87	80.28	73.36	71.54	72.69	80.28	80.05	76.57	74.51	73.82	77.22
CA1	71.53	74.84	85.86	87.47	87.14	70.74	74.48	87.80	98.14	89.33	84.94	78.59	76.87	79.58	76.33
CA2	72.53	85.21	91.78	86.50	86.50	87.29	87.48	73.97	86.82	87.29	87.14	73.14	82.00	77.68	76.48
CA3	72.53	85.21	91.78	86.50	86.50	87.29	87.48	73.97	86.82	87.29	87.14	73.14	82.00	77.68	76.48
CA4	76.97	84.72	84.77	79.74	100.0	83.62	81.63	82.52	75.53	87.76	88.92	86.31	91.60	75.08	79.58
CA5	76.97	84.72	84.77	79.74	100.0	83.62	81.63	82.52	75.53	87.76	88.92	86.31	91.60	75.08	79.58
epv1	73.61	79.92	87.32	74.72	86.81	86.81	88.16	89.16	76.33	88.86	100.0	91.18	86.38	80.00	78.74
epv2	73.61	79.92	87.32	74.72	86.81	86.81	88.16	89.16	76.33	88.86	100.0	91.18	86.38	80.00	78.74
IP90	71.61	75.69	74.03	72.61	74.52	74.52	77.73	74.54	73.34	75.34	76.37	75.73	81.76	87.36	83.08
LP4	71.61	75.69	74.03	72.61	74.52	74.52	77.73	74.54	73.34	75.34	76.37	75.73	81.76	87.36	83.08
LP8	71.61	75.69	74.03	72.61	74.52	74.52	77.73	74.54	73.34	75.34	76.37	75.73	81.76	87.36	83.08
ospE	76.39	87.50	85.92	85.42	89.58	89.58	85.18	77.08	86.11	89.58	85.18	86.12	78.35	77.80	81.68
p21	74.13	83.80	98.70	92.99	83.44	83.44	85.92	90.21	72.08	89.51	89.52	84.51	77.36	80.94	77.36
V5116	71.25	74.65	73.69	75.66	72.37	72.37	76.76	78.17	84.21	76.76	76.76	73.68	84.83	84.96	81.64
IKAc23	61.02	74.65	73.69	75.66	72.37	72.37	76.76	78.17	84.21	76.76	76.76	73.68	84.83	84.96	81.64
FGJ	58.82	67.83	71.52	70.13	69.54	69.54	69.93	68.21	79.57	68.21	68.21	73.43	64.24	75.33	93.92
Pdau	68.99	62.68	62.68	61.69	65.10	65.10	67.91	69.91	67.91	69.91	67.91	70.20	60.26	65.27	87.82

\* Percent nucleotide identity values are presented in the upper half of the table, and percent amino acid similarity values are given in the lower half of the table. The *p21* and *erp* sequences analyzed were previously determined by others; their accession numbers are provided in Table 1.

scribed *ospE*-related genes. As a consequence, most of the amplified genes could not be assigned with certainty a gene name designation. It appears instead that several genes in the *ospE* subfamily of the UHB gene family are in a somewhat gray area regarding nomenclature. Rather than assign a new gene name to each, and in an attempt to simplify the somewhat complicated nomenclature of *ospE*-related genes, we refer to them simply as *ospE* paralogs.

Alignment of the nucleotide sequences revealed a concentration of sequence variation between alignment positions 220 through 360 (variable domain 1). This domain is characterized by the presence of in-frame insertions and deletions, flanked in many isolates by direct repeat elements (Fig. 3). Some of the insertions are identical to those found in the *B. burgdorferi* 3. Relative to *B. burgdorferi* N40 *ospE*, the *B. andersonii* 210 *ospE* amplicon sequence carries an insertion of 102 nucleotides (nt) flanked by a 18-bp repeat containing three mismatches. At the same site in the CA4 and CA9 amplicons there is a 38-bp near-perfect repeat (37 of 38 nt) separated by 1 nt. In the IP90 amplicon this site carries a 34-bp repeat (32 of 34 nt) separated by 2 nt. A second interesting feature of the IP90 amplicon is a 12-bp insertion in the rest of the sequence coding for sequence which results in the presence of two potential ribosomal binding sites.

A second variable region occurs between positions 505 and 535 (variable domain 2). Diversity within this domain is significant yet is less pronounced than that seen in variable domain 1. At this site the *B. burgdorferi* LP3-derived amplicon carries a 24-nt insertion (relative to *osPE* of *B. burgdorferi* N40) identical in sequence to the insertion present in *erPC* of *B. burgdorferi* B31T. The *B. burgdorferi* CA3 and N40 amplicon sequences contain a 13-nt insertion at this position. *A. afzelii* PGau and *B. garinii* IP90 carry a similar 24-nt insertion (63% identical with each other) that is not related to the insertion observed in LP3. *B. anderssoni* 21038 harbors a unique 12-bp insertion at this site. In contrast to domain 1, domain 2 is not flanked by repeat elements.

Sequence analysis of the amplicons that were hybridization negative with the various *ospE* paralog-targeting probes (as

described above) revealed these sequences to be among the most divergent of the *ospE* paralogs. Near-complete sequences for 297, 272, and R100 amplicons and a complete sequence for the JDI amplicon were determined. These sequences were found to be identical to each other, and in light of this, we focus our discussion on the amplicon sequence from *B. burgdorferi* JDI. Of all *ospE*-related genes characterized thus far, this variant is most peripheral to the group in terms of sequence conservation. Its nucleotide identity with other *ospE* paralogs ranges from 67.25 to 84.40%. Like other *ospE*-related genes, it is polymorphic within variable domain 1. Relative to *B. burgdorferi* N40 *ospE*, it carries an insertion of 122 nt which is located just upstream by a single-base insertion that restores the reading frame. Interestingly, this insertion is at the insertion exhibit 90% identity with the insertion present in the *B. andersonii* 21038 amplicon sequence. Particularly notable is the homology between the 5' end of the insertion (and a small segment of its 5' flanking sequence) and another member of the UHB gene family, *erpK*. The gene from which this amplicon was derived appears to be a composite gene carrying segments of at least two UHB gene family members.

To compare the putative proteins encoded by the *ospE* amplicon sequences, the partial nucleotide sequences were translated and aligned (Fig. 4). To assess the relatedness of the translated sequences, pairwise sequence comparisons were performed (Table 4). These values ranged from 58.8% similarities, indicating that all the amino acid sequences are members of a single gene family. Analysis of the deduced amino acid sequences revealed several noteworthy features. All carry a relatively conserved, hydrophobic putative leader peptide of 15 to 20 amino acids and an amino acid motif similar to the proposed signal peptidase II site of *ospC* (LGAG) (14). Just downstream of the putative leader peptide within variable domain I of some isolates are amino acid repeat motifs of various lengths and sequences. It should be noted that not all sequences carry repeats at this nucleotide level exhibit repeats at the amino acid level (Fig. 4). In the IPD, the SLSDOG is perfectly and tandemly repeated four times. To assess the physical properties of the deduced amino acid se-

[illegible]

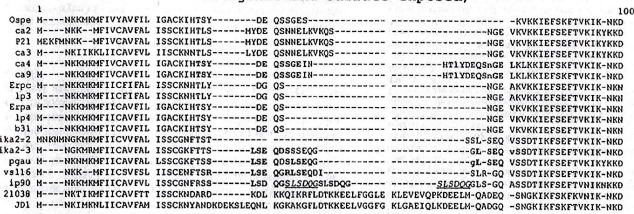
FIG. 3. Alignment of the  $\psi$ h(+)-E470 amplicon nucleotide sequences. Sequences were aligned by using the PILEUP program with some manual adjustment. The isolates analyzed are indicated at the left. For comparative purposes some previously determined gene sequences (*ospE*, *crpA*, *crpA*, and *p21*) were included, and these are indicated by their designated gene names (for accession numbers, see Table 1). Gaps are indicated by dashes. The binding sites of some oligonucleotide probes and primers (indicated above the alignment) are in boldface. Repeat elements are indicated by underlining, and mismatches are indicated by capital letters. Putative ribosomal binding sites (RBS) and translational start codons are also indicated. Since some of these sequences are partial and lack their 3' terminal, stop codons are not shown. The two different sequences from *B. japonica* IkA2 are the sequences determined for each of the two amplicons (IkA2-2 and IkA2-3) obtained from this isolate.







Variable domain 1 (hydrophilic, antigenic and surface exposed)



## Variable domain 2

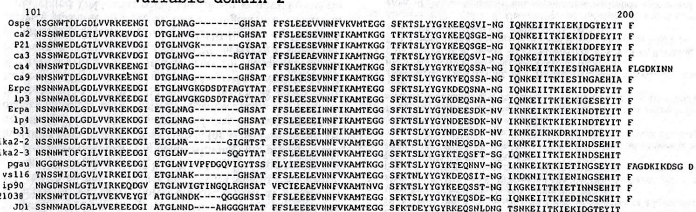


FIG. 5. Computer analysis of the possible physical properties of the deduced amino acid sequence of the *B. burgdorferi* JD1 *ospE*-related amplicon. The output from analyses conducted by using the Genetics Computer Group, Wisconsin, Sequence Analysis Package, version 9.0, as described in the text, is shown. In all cases default values were used. Variable domain 1 spans residues 20 through 75. KD, Kyte-Doolittle; Prob., probability.

variable domain 1 predict them to be hydrophilic and antigenic. Surface exposure of this region is predicted for all sequences, and this is consistent with the fact that these are

these surface-exposed proteins, and this in turn could influence the host-pathogen interaction.

The relatively high sequence identities of the *ubd3(+)* and *E470(-)* amplicon sequences with *ospE*, *erpA*, *erpC*, and *p21* imply that all of these genes are derived from a common ancestral gene. Most of the sequences determined here could not be definitively assigned any of the previously proposed

FIG. 4. Alignment of the deduced amino acid sequences. The amino acid alignment was generated as described in the text with some manual adjustment. The isolate from which a sequence was obtained is indicated at the left. Some previously determined sequences (6, 8, 20, 25) were included in the alignment, and the corresponding protein names are indicated on the left. Repeat motifs are indicated by boldface. In cases where the repeats are tandem, alternating copies are indicated by italics and underlining. Mismatches among the repeat elements are indicated by lowercase letters.

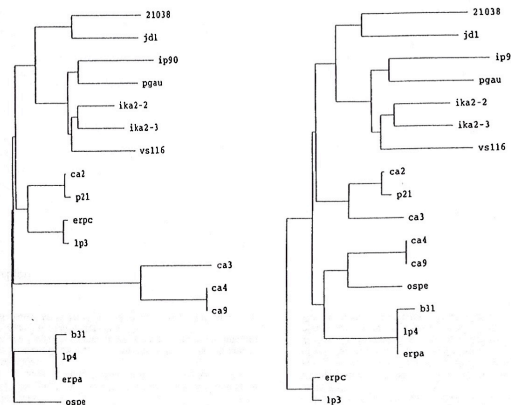
branch lengths. This observation indicates that the variable domains are not evolutionarily stable. Hence, while all of the analyzed sequences are clearly derived from a common ancestral gene, it is evident that the variable domains have been influenced by recent molecular events. Hence, the overall gene trees have been influenced by events that appear to have included rearrangements and recombination (i.e., insertions, deletions, and/or gene fusions). Consistent with this, we previously provided evidence for the existence of gene fusions between *ospE* and *ospF* in some *B. burgdorferi* isolates (20).

## DISCUSSION

In light of the putative role of members of the UHB gene family in adaptation to the mammalian environment or in virulence (1, 8, 26), an assessment of their genetic variability and distribution is an important step towards defining their potential biological role and true biological significance. Here we demonstrate that while the UHB gene family is universal among *B. burgdorferi* sensu lato complex species, there is significant variation in its composition and organization at both the inter- and intraspecies levels as evidenced by the observed RFLP patterns and polymorphic UHB gene family-derived PCR amplicons. The UHB RFLP patterns do not reflect nat-

terms of phylogenetic divergence in the *B. burgdorferi* sensu lato complex (2, 3, 15–17), suggesting that this gene family is evolutionarily unstable and that recombination in and among these genes has influenced their sequence and organization. As a result, the organization of this gene family as originally described for *B. burgdorferi* B31T (6, 25) may not serve as a universal model for its organization in other isolates.

The RFLP pattern and PCR analyses demonstrated that members of the *aspE* subfamily of the UHB gene family are polymorphic, and to identify the molecular basis for these polymorphisms, DNA sequence analyses were performed. Although the uidY<sup>+</sup>-E470(-) amplicons exhibit relatively high sequence identity values, defined domains of variability were identified. Variable domain 1 is characterized by the presence of variable-length direct repeat elements (some as long as 38 bp) and insertions (relative to the *B. burgdorferi* N40 *aspE* sequence). In light of the clustering of these repeats around sequence 1, it is conceivable that the repeats have contributed to recent molecular rearrangements in variable domain 1. It was obtained through the construction of phylograms. Computer analyses of the deduced amino acid sequences within





gene name designations (i.e., *erpA*, *erpC*, *erpI*, and *p21*), since they are evolutionarily equidistant from each gene. In light of this, it is unclear if the designations that have been assigned to *ospE*-related genes are useful. While the cp32s of *B. burgdorferi* B31T, which carry some *ospE*-related genes, can be differentiated through complex restriction analyses and subsequent Southern blotting (6), the data presented here demonstrate that differentiation of the *ospE* paralogs themselves in other isolate populations would not be straightforward. In view of these considerations, we have opted not to assign new gene name designations for the novel *ospE*-related genes described in this report and instead refer to them only as *ospE* paralogs. The analyses presented here strongly suggest that *ospE*, *p21*, *erpA*, *erpC*, and *erpI* are genetically synonymous. In cases where the full complement of *ospE* paralogs within an isogenic isolate are defined, numerical qualifiers could prove useful for their differentiation, e.g., *ospE1* and *ospE2*, etc., as has been done for the *B. burgdorferi* vls genes (28).

The specific molecular mechanisms by which genetic diversity is generated in *ospE*-related genes remain to be determined. The presence of direct repeat elements in many of the sequences flanking variable domain 1 is intriguing. It is conceivable that diversity in defined domains of *ospE*-related genes could arise from transposition, mutation, or recombination or perhaps through the exchange of sequence cassettes. Suggestive evidence for gene mosaicism can be found in the sequence of the *ospE*-related amplicon from *B. burgdorferi* JD1, which harbors sequence blocks with high identity to segments of *B. burgdorferi* B31T *erpC* and the insertion element of the *B. andersonii* 21038 amplicon. Although the pressures that may drive putative gene mosaicism in the UHB gene family are undefined, it has been demonstrated that the exchange of sequence cassettes among related sequences plays a key role in the generation of antigenic diversity in the vls genes of *B. burgdorferi* (28) and the pil genes of *Neisseria gonorrhoeae* (10). We have previously presented evidence for the lateral transfer of plasmids among species of the *B. burgdorferi* sensu lato complex (19). If lateral transfer of UHB-flanked *ospE* variant-carrying plasmids also occurs, this could generate a source of *ospE*-related DNA that could participate in homologous recombination increasing the potential number of possible *ospE* paralogs. While evidence for direct recombination in the UHB gene family during *in vitro* cultivation has not yet been demonstrated, the data presented here suggest that in natural populations recombination or mutation has occurred and has been a contributing factor to the organization, composition, and sequence of the UHB gene family. Long-term studies are underway to assess the frequency of recombination or mutation in the *ospE* subfamily in isogenic clones of *B. burgdorferi* during infection in mice. The significance of mutation and/or recombination in these genes and its influence on the biology of these important pathogens represents the next important area to be assessed in the study of the UHB gene family.

#### ACKNOWLEDGMENTS

We acknowledge our colleagues in the Molecular Pathogenesis group at Virginia Commonwealth University for their input and helpful discussions and the many individuals who contributed isolates which have made these analyses possible.

This research was supported in part by grants from the National Institutes of Health (5R29AI37787) and the Jeffress Trust.

#### REFERENCES

- Akins, D. S., F. Porcella, T. G. Popova, D. Shevchenko, S. I. Baker, M. Li, M. V. Norgard, and J. D. Radolf. 1995. Evidence for *in vivo* but not *in vitro* expression of a *Borrelia burgdorferi* outer surface protein F (OspF) homologue. *Mol. Microbiol.* 18:507-520.
- Assou, M. V., D. Postic, G. Paul, P. Nevel, and G. Baranton. 1994. Identification of two new genomic groups among American *Borrelia burgdorferi* sensu lato strains. *FEMS Microbiol. Lett.* 121:93-98.
- Baranton, G., D. Postic, I. Saint Girons, P. Borrelli, J.-C. Piffaretti, M. Assou, and P. A. D. Grimont. 1992. Delineation of *Borrelia burgdorferi* sensu stricto, *Borrelia garinii* sp. nov., and group VS461 associated with Lyme borreliosis. *Int. J. Syst. Bacteriol.* 42:378-383.
- Barbour, A. G., C. J. Carter, V. Bundorf, and J. Hinnebusch. 1996. The nucleotide sequence of a 1.6 kb plasmid of *Borrelia burgdorferi* reveals similarities to those of circular plasmids of other prokaryotes. *J. Bacteriol.* 178:6635-6639.
- Carlson, J. A., C. LaVoie, S. Y. Sung, and R. T. Marconi. 1998. Analysis of the organization of multiplicity linear and circular plasmid-carried open reading frames in *Borrelia burgdorferi* sensu lato isolates. *Infect. Immun.* 66:1149-1158.
- Casjens, S., R. van Yugi, K. Tilly, P. A. Rosa, and B. Stevenson. 1997. Homology throughout the multiple 32-kilobase circular plasmids in Lyme disease spirochetes. *J. Bacteriol.* 179:217-227.
- Champion, C. I., D. R. Blanco, J. T. Skare, D. A. Haake, M. Giladi, D. Foley, J. N. Miller, and M. A. Lovett. 1994. A 9.8-kilobase pair circular plasmid of *Borrelia burgdorferi* encodes an expected protein: evidence for expression only during infection. *Infect. Immun.* 62:2653-2661.
- Das, S., S. W. Barthold, S. S. Giles, R. R. Montgomery, S. R. Telford III, and E. Fikrig. 1997. Temporal pattern of *Borrelia burgdorferi* p21 expression in ticks and the mammalian host. *J. Clin. Invest.* 99:907-915.
- Fraser, C., 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. Fleischman, D. Richardson, J. Peterson, A. R. Kerlavage, J. Quackenbush, S. Salzberg, M. Hanson, R. Yagi, N. Palmer, M. D. Adams, J. Geocayne, J. Weidman, T. Utterback, L. Wathley, L. McDonald, P. Artlich, C. Bowman, S. Garland, F. C. M. C. Cotton, K. Horst, K. Roberts, B. Hatch, B. O. Smith, and J. C. Venter. 1997. Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature* 390:580-586.
- Hagblom, P., E. Segal, E. Billiard, and M. So. 1985. Intragenic recombination leads to leads to plus antigenic variation in *Neisseria gonorrhoeae*. *Proc. Natl. Acad. Sci. USA* 81:155-158.
- Hinnebusch, J., and A. G. Barbour. 1991. Linear plasmids of *Borrelia burgdorferi* have a telomeric structure and sequence similar to those of a eukaryotic virus. *J. Bacteriol.* 173:7233-7239.
- Hupp, T. P., and K. R. Woods. 1981. Prediction of protein antigenic determinants from amino acid sequences. *Proc. Natl. Acad. Sci. USA* 78:3824-3828.
- Janesson, B. A., and H. Wolf. 1988. The antigenic index: a novel algorithm for predicting antigenic determinants. *CABIOS* 4:181-186.
- Lam, T. T., T.-P. K. Nguyen, R. R. Montgomery, F. S. Kantor, E. Fikrig, and R. A. Flavell. 1994. Outer surface proteins E and F of *Borrelia burgdorferi*, the agent of Lyme disease. *Infect. Immun.* 62:289-298.
- Marconi, R. T., and C. F. Garon. 1992. Identification of a third genomic group of *Borrelia burgdorferi* through signature nucleotide analysis and 16S rRNA sequence determination. *J. Gen. Microbiol.* 138:533-536.
- Marconi, R. T., and C. F. Garon. 1992. Phylogenetic analysis of the genus *Borrelia*: a comparison of North American and European isolates of *B. burgdorferi*. *J. Bacteriol.* 174:241-244.
- Marconi, R. T., D. Liveris, and I. Schwartz. 1995. Identification of novel insertion elements, restriction fragment length polymorphism patterns, and discontinuous 23S rRNA in Lyme disease spirochetes: phylogenetic analyses of rDNA genes and their intergenic spaces in *Borrelia japonica* sp. nov. and genomic group 21038 (*Borrelia andersonii* sp. nov.) isolates. *J. Clin. Microbiol.* 33:2427-2434.
- Marconi, R. T., D. S. Samuels, and C. F. Garon. 1993. Transcriptional analysis and mapping of the *ospC* gene in Lyme disease spirochetes. *J. Bacteriol.* 175:926-932.
- Marconi, R. T., D. S. Samuels, R. K. Landry, and C. F. Garon. 1994. Analysis of the distribution and molecular heterogeneity of the *ospD* gene among the Lyme disease spirochetes: evidence for lateral gene exchange. *J. Bacteriol.* 176:4572-4582.
- Marconi, R. T., S. Y. Sung, C. N. Hughes, and J. A. Carlson. 1996. Molecular and evolutionary analyses of a variable series of genes in *Borrelia burgdorferi* that are related to *ospE* and *ospF*, constitute a gene family, and share a common upstream homology box. *J. Bacteriol.* 178:5615-5626.
- Marconi, R. T., J. Wigboldus, H. Weisback, and N. Brot. 1991. Transcription start site and *MitR* binding sites on the *Escherichia coli* met H gene. *Biochem. Biophys. Res. Commun.* 175:1057-1063.
- Porcella, F., T. G. Popova, D. R. Akins, M. Li, J. R. Radolf, and M. V. Norgard. 1996. *Borrelia burgdorferi* supercoiled plasmids encode multiplicity open reading frames and a lipoprotein gene family. *J. Bacteriol.* 178:3293-3307.
- Schwartz, J. J., A. Gaszymian, and I. Schwartz. 1992. rRNA gene organization in the Lyme disease spirochaete, *Borrelia burgdorferi*. *J. Bacteriol.* 174:3757-3765.
- Simpson, W. J., C. F. Garon, and T. G. Schwan. 1990. Analysis of supercoiled circular plasmids in infectious and non-infectious *Borrelia burgdorferi*. *Microb. Pathog.* 8:109-118.
- Stevenson, B., K. Tilly, and P. A. Rosa. 1996. A family of genes located on four separate 32-kilobase circular plasmids in *Borrelia burgdorferi* B31. *J. Bacteriol.* 178:3508-3516.
- Suk, K., S. Das, W. Sun, B. J. Wang, S. W. Barthold, R. A. Flavell, and E. Fikrig. 1995. *Borrelia burgdorferi* genes selectively expressed in the infected host. *Proc. Natl. Acad. Sci. USA* 92:4269-4273.
- Wallach, R., C. Brenner, M. D. Kramer, and M. M. Simon. 1995. Molecular cloning and immunological characterization of a novel linear-plasmid-encoded gene, *pG*, of *Borrelia burgdorferi* expressed only *in vivo*. *Infect. Immun.* 63:3327-3335.
- Zhang, J.-R., J. M. Hardham, A. G. Barbour, and A. G. Norris. 1997. Antigenic variation in Lyme disease borreliae by promiscuous recombination of vmp like sequence cassettes. *Cell* 89:275-285.
- Zuckert, W. R., and J. Meyer. 1986. Circular and linear plasmids of Lyme disease spirochetes have extensive homology: characterization of a repeated DNA element. *J. Bacteriol.* 178:2287-2298.

Editor: J. T. Barbieri