

## Molecular Cloning and Characterization of the *Ehrlichia chaffeensis* Variable-Length PCR Target: an Antigen-Expressing Gene That Exhibits Interstrain Variation

JOHN W. SUMNER,\* JAMES E. CHILDS, AND CHRISTOPHER D. PADDOCK

*Viral and Rickettsial Zoonoses Branch, Division of Viral and Rickettsial Diseases, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333*

Received 27 October 1998/Returned for modification 17 December 1998/Accepted 5 February 1999

A clone expressing an immunoreactive protein with an apparent molecular mass of 44 kDa was selected from an *Ehrlichia chaffeensis* Arkansas genomic library by probing with anti-*E. chaffeensis* hyperimmune mouse ascitic fluid. Nucleotide sequencing revealed an open reading frame (ORF) capable of encoding a 198-amino-acid polypeptide. The ORF contained four imperfect, direct, tandem 90-bp repeats. The nucleotide and deduced amino acid sequences did not show close homologies to entries in the molecular databases. PCR with primers whose sequences matched the sequences flanking the ORF was performed with DNA samples extracted from cell cultures infected with nine different isolates of *E. chaffeensis*, blood samples from seven patients with monocytic ehrlichiosis, and *Amblyomma americanum* ticks collected in four different states. The resulting amplicons varied in length, containing three to six repeat units. This gene, designated the variable-length PCR target, is useful for PCR detection of *E. chaffeensis* and differentiation of isolates.

Human ehrlichiosis was first described in the United States in 1987 (17), and the first isolation of an *Ehrlichia* species from a North American patient was reported in 1991 (8). The organism was designated a new species, *Ehrlichia chaffeensis*, because the 16S rRNA gene sequence of the isolate was significantly different from that of previously described *Ehrlichia* species (2). Infection with *E. chaffeensis* results in a moderate to severe febrile illness observed most frequently in the southeastern and south-central regions of the United States (11, 13, 28). Over 70% of patients describe a history of tick bite or exposure 1 to 2 weeks preceding the illness. Several studies have implicated the lone star tick, *Amblyomma americanum*, as the primary vector of *E. chaffeensis* (4, 12, 16). Approximately 750 cases of *E. chaffeensis* infection have been confirmed serologically by the Centers for Disease Control and Prevention (CDC) to date (5).

The emergence or recognition of human diseases caused by tick-transmitted ehrlichiae has stimulated interest in the molecular biology of these obligate intracellular bacteria. The cloning and characterization of several antigen-expressing genes belonging to *E. chaffeensis* have recently been described, including the *groESL* heat shock operon (24, 25) and the 120-kDa immunodominant surface protein gene (29, 30). Ohashi et al. (18) and Reddy et al. (22) have described a multigene family of *E. chaffeensis* that demonstrates homology to major surface antigen genes (MAP 1) of a closely related bacterium, *Cowdria ruminantium* (21, 27).

Initially, *E. chaffeensis* was isolated and propagated in cell culture with difficulty, but more recently, multiple isolations of this pathogen from human patients from several geographic areas have been described. The original isolate was designated the Arkansas strain (2, 8). The 91HE17 and Sapulpa isolates were described in 1995 (10) and 1997 (6), respectively. Since 1996, six additional isolates of *E. chaffeensis* have been ob-

tained at CDC from blood samples from patients with ehrlichiosis (reference 19 and data herein). The availability of multiple isolates now provides ample material for the study of the molecular and immunologic diversity among different strains of *E. chaffeensis*. In this report, we describe the cloning and sequencing of an *E. chaffeensis* gene containing repetitive sequence motifs. We have referred to this gene as the variable-length PCR target (VLPT) and have described sequence differences among PCR products amplified from the Arkansas, Jax, and St. Vincent isolates of *E. chaffeensis* (19). In this report, we describe the detection of VLPT by PCR in a variety of samples from different sources, sequence variation among the amplicons, and recombinant expression of the VLPT protein in *Escherichia coli*.

### MATERIALS AND METHODS

**Isolation and cultivation of *E. chaffeensis*.** The Arkansas, 91HE17, Sapulpa, Jax, and St. Vincent isolates of *E. chaffeensis* were obtained from clinical samples as described previously (6, 8, 10, 19). The Liberty, Osceola, Wakulla, and West Paces isolates were recovered from EDTA-anticoagulated whole blood collected from patients with clinical disease by a previously described method (19). Briefly, patient blood samples (1 to 5 ml) were diluted with 2 volumes of sterile Hanks' balanced salt solution and layered onto Histopaque 1083 (Sigma Diagnostics, St. Louis, Mo.). The prepared samples were centrifuged at 400 × g for 10 min. The leukocyte pellets were resuspended in culture medium (minimal essential medium containing Earle's salts and 16 mM sodium bicarbonate [GIBCO BRL, Grand Island, N.Y.], 8.8% heat-inactivated fetal bovine serum [Hyclone Laboratories, Logan, Utah], 1.8 mM L-glutamine [GIBCO BRL], 0.1 mM minimal essential medium with nonessential amino acids [GIBCO BRL], and 8.8 mM HEPES buffer [GIBCO BRL]). Semiconfluent monolayers of DH82 cells in a 25-cm<sup>2</sup> polystyrene cell culture flask were inoculated with the cell suspensions. Cultures were incubated at 37°C in a 5.0% CO<sub>2</sub> atmosphere. Isolates of *E. chaffeensis* were maintained in continuous culture in DH82 cells as described previously (8, 19).

**Construction of genomic expression library.** Cell-free preparations of *E. chaffeensis* Arkansas were obtained from infected DH82 cell cultures as described previously (24). DNA was extracted by proteinase K digestion and phenol-chloroform extraction and was fragmented for library construction by using *EcoRI* star activity (1). DNA fragments from 2 to 10 kb in size were isolated by agarose gel electrophoresis and were cloned into Stratagene's Lambda Zap II vector by the procedures recommended by the manufacturer (Stratagene, La Jolla, Calif.).

**Immunoscreening of expression library.** Clones were screened by using anti-*E. chaffeensis* hyperimmune mouse ascitic fluid (HMAF; kindly provided by Tom Ksiazek, Special Pathogens Branch, CDC). Preparation of HMAF has been

\* Corresponding author. Mailing address: Viral and Rickettsial Zoonoses Branch, Centers for Disease Control and Prevention, 1600 Clifton Rd., Mailstop G-13, Atlanta, GA 30333. Phone: (404) 639-1075. Fax: (404) 639-4436. E-mail: jws3@cdc.gov.

described previously (14). Recombinant clones expressing reactive proteins were selected by immunoscreening phage plaques on lawns of *E. coli* XL-1 Blue. Nitrocellulose filters soaked in 10 mM isopropyl- $\beta$ -D-thiogalactopyranoside and placed over lawns during plaque formation were reacted with a 1:200 dilution of HMAF for 1 h in Tris-buffered saline containing Tween 20 (TBST). The filters were washed three times in TBST and reacted for 1 h with anti-mouse immunoglobulin G (heavy and light chains) peroxidase conjugate (Kirkegaard & Perry Laboratories, Gaithersburg, Md.) diluted 1:5,000 in TBST. The filters were then washed three times with TBST and reacted with tetramethylbenzidine membrane substrate (Kirkegaard & Perry). Immunoreactive plaques were subjected to three rounds of plaque purification. Phagemids from reactive plaques were rescued and excised by the Stratagene protocol, producing pBluescript plasmids containing inserts.

**Immunoblotting.** *E. coli* containing excised recombinant phagemid or transformed with plasmid subclones was grown overnight at 37°C in 5 ml of Luria broth containing ampicillin (100  $\mu$ g/ml). Fresh cultures containing 1 mM isopropyl- $\beta$ -D-thiogalactopyranoside were subsequently inoculated with 50  $\mu$ l of the overnight culture. Several 200- $\mu$ l samples were removed when the optical density of the culture reached 0.3 to 0.4 (600-nm wavelength), and the bacterial cells were pelleted by centrifugation at 10,000  $\times$  g for 3 min. The bacteria were resuspended in sample buffer (Novex, San Diego, Calif.) containing 5% 2-mercaptoethanol and 1% sodium dodecyl sulfate (SDS), heated at 100°C for 5 min, and centrifuged at 10,000  $\times$  g for 3 min. Ten-microliter samples of each supernatant were loaded into individual wells of 4 to 20% gradient Tris-glycine-SDS gels (Novex) for SDS-polyacrylamide gel electrophoresis at 100 V for 90 min. The proteins were transferred to nitrocellulose membranes (Novex) for 2 h at 90 V with a Mini-Transblot Cell (Bio-Rad, Hercules, Calif.). The filters were reacted with the HMAF as described above for immunoscreening.

**Subcloning.** The restriction endonucleases, calf intestinal alkaline phosphatase, and T4 DNA ligase used for mapping and the construction of subclones were obtained from Boehringer Mannheim, Indianapolis, Ind., or New England Biolabs, Beverly, Mass. Some subclones were made by cloning PCR products directly into the T/A cloning vector pGEM-T (Promega Corp., Madison, Wis.) according to the manufacturer's recommendations.

**Nucleotide sequencing.** Plasmids were purified with Wizard Minipreps (Promega Corp., Madison, Wis.), and Wizard PCR preps (Promega Corp.) were used for purification of PCR products. PCR products were sequenced directly, with the exception of one product obtained from an individual tick (tick 97-36; see Table 1) that was cloned into plasmid pGEM-T prior to sequencing (Promega Corp.). If gel purification was necessary prior to sequencing, 40  $\mu$ l of the individual PCR mixture was electrophoresed in gels containing 1.2% low-melting-point agarose (Boehringer Mannheim) and the appropriate band was excised. Purified preparations were sequenced with the Prism Ready Reaction Dye/Deoxy Cycle Sequencing kit (Applied Biosystems, Foster City, Calif.) and a Perkin-Elmer 9600 thermocycler. Thermocycler parameters for sequencing were 96°C for 1 min, 50°C for 15 s, and 60°C for 4 min for 25 cycles. Unincorporated fluorescence-labeled deoxynucleoside triphosphates were removed with Centr-Sep columns according to the manufacturer's recommendations (Princeton Separations, Inc., Adelphi, N.J.). Samples were loaded onto 5% polyacrylamide gels for electrophoresis and detection on Applied Biosystem model 370A or 377 automated sequencers. Both strands were sequenced, initially by primer walking and later with established primer sets.

**Samples selected for analysis of the VLPT gene.** Three categories of specimens were evaluated by PCR for detection of the VLPT gene (see Table 1): (i) EDTA-anticoagulated whole-blood samples from patients with *E. chaffeensis* infection confirmed by PCR with primers HE1 and HE3 (3) for detection of the 16S rRNA gene; (ii) nine isolates of *E. chaffeensis*, in DH82 cells, obtained from human patients (Arkansas, 91HE17, and Sapulpa [kindly provided by D. H. Walker], Jax, St. Vincent, Osceola, Wakulla, Liberty, and West Paces); and (iii) individual and pooled *A. americanum* ticks collected from geographic regions where *E. chaffeensis* is endemic.

**Extraction of *E. chaffeensis* DNA from patient blood, isolates, and ticks.** DNA was extracted from 200  $\mu$ l of patient whole blood by using the QIAamp Blood Kit (Qiagen Inc., Santa Clarita, Calif.) according to the manufacturer's recommendations. Similarly, DNA was extracted from 20  $\mu$ l of supernatant from *E. chaffeensis*-infected DH82 cell cultures. Individual ticks were minced in a cryotube with a sterile scalpel blade, and DNA was extracted with QIAamp tissue kits (Qiagen, Inc.). Extraction of DNA from pooled *A. americanum* ticks (see Table 2) was conducted at CDC's Ft. Collins facility as part of a previous study, and the extraction procedure has been described previously (4, 15).

**PCR amplification.** PCRs were performed with GeneAmp kits (Roche Molecular Systems, Inc., Branchburg, N.J.) according to the manufacturer's recommendations and Perkin-Elmer thermal cyclers. Ten microliters from each DNA extraction was added to 90  $\mu$ l of the master mixture for each 100- $\mu$ l PCR mixture. The final reagent concentrations were 0.5  $\mu$ M each primer, 2.5 U of AmpliTaq DNA polymerase per reaction mixture, 10 mM Tris-HCl (pH 8.3), 50 mM KCl, 1.5 mM MgCl<sub>2</sub>, and each deoxynucleoside triphosphate at a concentration of 200  $\mu$ M. The following thermocycler parameters were used: 3 cycles of 94°C (1 min), 52°C (2 min), and 70°C (90 s), followed by 37 cycles of 88°C (1 min), 55°C (2 min), and 70°C (90 s), followed by an extension period (68°C, 5 min). VLPT primers FB5A (5'-GTGACATCTTAGTTTAATAGAAC) and FB3A (5'-AAGACTGAAACGTTATAGAG) were used in primary PCRs.

"Nested PCR with primers FB5 (5'-AAATAGGGTATAAATATGTCAC) and FB3 (5'-GCCTAATTGAGATAAATACTAAC) was performed with extracts from individual ticks. One microliter of the primary PCR product was used as a template in the nested reactions. The cycling parameters for the nested PCRs were the same as those used for the primary reactions. Samples were also tested by a PCR assay with primers HE1 and HE3 targeted to the 16S rRNA gene (rDNA) of *E. chaffeensis* as described by Anderson et al. (3). In addition, extracts from individual isolates in cell culture were tested by PCR with primers WF1 and WR2 for amplification of the 120-kDa antigen gene as described by Yu et al. (29). PCR products were detected by electrophoresis of 10- $\mu$ l samples in 1.4% Tris-acetate agarose gels containing ethidium bromide.

**Computer analysis of sequence data.** Nucleotide sequences were edited and assembled with the TED and XBP programs of the STADEN sequence analysis package (23). Nucleotide sequence homology searches were made through the National Center for Biotechnology Information BLAST network service. Sequence homology comparisons were made with the GAP program of the GCG package (Genetics Computer Group, Madison, Wis.). Multiple-sequence alignments were made with the PILEUP and PRETTY programs of the GCG package.

**Nucleotide sequence accession numbers.** The nucleotide sequence accession numbers for the complete VLPT sequences of the following *E. chaffeensis* isolates are as indicated: 91HE17, AF121237; Arkansas, AF121232; Jax, AF121234; Liberty, AF121236; Osceola, AF121233; Sapulpa, AF121230; St. Vincent, AF121231; Wakulla, AF121238; and West Paces, AF121235.

## RESULTS

**Sequence analysis of *E. chaffeensis* Arkansas clone.** pBluescript clones derived from 10 immunoreactive plaques were analyzed by SDS-polyacrylamide gel electrophoresis and immunoblotting with HMAF. Three of 10 clones expressed an immunoreactive protein that comigrated with an *E. chaffeensis* Arkansas protein with an apparent molecular mass of approximately 44 kDa. Restriction endonuclease digestion with *Eco*RI showed that each of the three clones contained an insert of approximately 7 kb, and nucleotide sequencing with M13 universal and reverse primers showed identical sequences for the inserts of each clone near the plasmid junction. One clone was selected for further analysis. After restriction mapping, subclones were made and screened for protein expression. A subclone containing a 2.6-kb fragment of the original insert expressed a 44-kDa protein. The complete insert was sequenced by primer walking. The longest open reading frame (ORF) encoded a 198-amino-acid polypeptide, approximately one-half the size predicted for a 44-kDa protein. A sequence compatible with a ribosome-binding site was found 11 bp upstream of the putative ATG start codon (Fig. 1). Multiple stop codons were found downstream of the first stop codon, and homologous sequences were not found elsewhere in the insert. Sequence homology searches with the nucleotide and deduced amino acid sequences did not show significant similarities to entries in the genetic databases.

**VLPT PCR.** Multiple sample types were evaluated by using the primers derived from the sequenced clone. Analysis of sequences from nine isolates, seven clinical whole-blood specimens, seven *A. americanum* tick pools, and one individual tick are listed in Table 1. Amplicons of four different sizes were produced, with each differing in apparent size by a factor of 90 bp (Fig. 2). Nucleotide sequencing revealed that the size differences resulted from variations in the number of repeat units. The corresponding patient blood samples from which six of the isolates (Jax, Liberty, Osceola, St. Vincent, Wakulla, and West Paces) were obtained were tested individually. For each blood sample, the VLPT sequence was identical to the VLPT sequence of the corresponding isolate (data not shown). The seven blood samples listed in Table 1 were collected from ehrlichiosis patients for whom isolates were not obtained.

In addition to demonstrating concordance between repeat unit number for isolates and the corresponding whole blood samples, we investigated whether the number of repeat units in the VLPT remained stable with passage in cell culture. DNA



-125 GTGACATCTTAGTTTATGAGAACTTTATATTTATATATGATTAAATATATA  
FB5A >  
-75 ATGATAATGGTATGTGGTTATACTGCTTATTAGTTGATCATGTACCTGT  
(RBS) M S Q F S E D N  
-25 GTGTTATGTTAAATAGGGTATAAATATGTCACAACTTCTCTGAAGATAATA  
FB5 >  
M G N I Q M P F D S D S H E P S H  
26 TGGGTAATATACAAATGCCTTTTGAATTCGATTACATGAGCCCTTCTCAT  
L E L P S L S E E V I Q L E S D L  
76 CTGAGCTACCTAGTCTTTCTGAAGAAGTGATTCAATTAGAGAGTGATCT  
Q Q S S N S D L H G S F S V E L  
126 ACAACAATCTTCTAATTCGATTACACGGGCTTTTCTGTTGAGTTAT  
F D P F K E A V Q L G N D L Q Q S  
176 TTGATCCCTTTTAAAGAAGCAGTTCATTTGGGGAATGATCTACAACATCT  
S D S D L H G S F S V E L F D P S  
226 TCTGATTCGATTACACGGGCTTTTCTGTTGAGTTATTGATCCTTC  
K E E V Q L E S D L Q Q S S N S  
276 TAAAGAAGAAGTTCATTTGGAGAGTGATCTACAACAATCTTCTAATCTG  
D L H E S S F V E L P G P S K E E  
326 ATTTCACGAGTCTCTCTTTTGTGAGTTACCTGGTCTTCCAAAGAAGAA  
V Q F E D D A K N V V Y G Q D H V  
376 GTTCAATTGGAAGATGATGCTAAAGATGATGATATGGAACAGACCATGT  
S L S E L G L L L G G V F S T M  
426 TAGTTTATCTGAATTAGGCTTATTGTTAGGTGGTGTTTTAGTACAATGA  
< FB3  
N Y L S G Y T P Y Y Y H H Y C C Y  
476 ATTATTGCTGCTGTTATACACCGTATTATTATCATCATTATTGTTGTTAT  
N P Y Y Y F D Y V T P D Y C H H C  
526 AATCCTTATTATTATTGATTATGTTACTCCAGATTATTGTCATCACTG  
S E S S L E \*  
576 TAGTGAAAGTAGTTTATAGTAGGATATTTGAATATATAAATGGTTGTTGA  
626 CTTCAAAAAGGTGTAGTTTATATGTTTATGCTGTTTATAGTGTAT  
676 AAGGATATGAGTTGTTTACTATTCTCTATAACGTTTCAGTCTT  
< FB3A

FIG. 1. Nucleotide and deduced amino acid sequences of the *E. chaffeensis* Arkansas VLPT. The following features are indicated in boldface type: putative ribosome-binding site (RBS), translation initiation codon, and termination codon (asterisk). An aspartic acid codon (GAT) and the sequence 5'-GTTTATAT, which are absent from sequences amplified from some strains of *E. chaffeensis*, and four individual nucleotide positions where substitutions (A or G) occur among different strains are also indicated in boldface type. PCR primers are underlined and labeled, and directions are indicated. The nucleotide sequence was numbered by designating the A of the putative translation initiation codon number 1.

samples were extracted from the St. Vincent, Jax, Osceola, and Wakulla isolates after the eighth cell culture passage. The sizes of the amplicons obtained from patient blood samples were identical to those obtained from the corresponding isolate after the eighth cell culture passage, indicating that the number of repeat units remained stable through multiple passages in cell culture.

**PCR specificity.** DNA samples extracted from the blood of a healthy patient and uninfected DH82 cells were included as negative controls and did not produce amplicons. DNA samples extracted from related bacteria were tested in PCRs with VLPT primers (FB5A-FB3A and FB5-FB3) to assess primer specificity. Samples included DNA extracted from cultures individually infected with the human granulocytic ehrlichiosis agent (strain USG3), *Ehrlichia canis* Florida, *Ehrlichia muris* AS145, and *Ehrlichia risticii* HRC-IL and from purified *Anaplasma marginale* Virginia, *Bartonella henselae* Houston 1, and *Rickettsia rickettsii* R. DNA extracted from *E. coli* XL-1 Blue, which was used for construction of the genomic library, was also tested. None of the DNA samples produced bands of

significant intensity and size compatible with those of the VLPT amplicons obtained from *E. chaffeensis*. DNA extracted from whole blood from a dog naturally infected with *Ehrlichia ewingii* produced an amplicon of approximately 1,600 bp in a primary PCR with primers FB5A and FB3A. This amplicon was sequenced and did not contain an ORF with sequence homology to VLPT. DNA samples extracted from several strains of *C. ruminantium* were tested with VLPT primers FB5 and FB3, and amplicons of different sizes were obtained from different strains; however, none of the products hybridized with a labeled *E. chaffeensis* VLPT probe (20a).

**Detection of *E. chaffeensis* in ticks.** To determine whether VLPT PCR would be useful for detection of *E. chaffeensis* in ticks and to determine whether a variable number of repeat units would also be present in samples of nonhuman origin, 10 pools of *A. americanum* ticks were tested with primers FB5A and FB3A. Seven of the pools were selected on the basis of previous positive results for the *E. chaffeensis* 16S rDNA (4) and three pools that produced negative results were selected as controls (Table 2). The VLPT PCR results (Table 2) showed an exact correlation with the 16S rDNA PCR results. Three VLPT amplicon sizes were obtained, and these corresponded to three, four, or five repeat units (Table 1). Two bands were visible in the PCR product from pool 95: a bright band comigrating with a three-repeat unit marker and a faint band comigrating with a four-repeat unit marker (data not shown). We were not able to recover sufficient DNA from the faint band for sequencing, but these results may indicate that more than one variant was detected in the pool.

Nested VLPT PCR was performed with DNA extracted from 42 individual adult *A. americanum* ticks collected in Baker County, Fla. Negative controls, a water extraction blank, and a colony-raised adult female *A. americanum* tick did not produce products. As a positive control, an adult female *A. americanum* tick infected with *E. chaffeensis* Arkansas by capillary feeding (20) produced a four-repeat-unit amplicon, as expected. Six of 42 tick extracts produced amplicons of the VLPT four-repeat-unit size. One of the six amplicons was sequenced and was confirmed to be VLPT (Table 1). Primary PCR with VLPT primer pair FB5A-FB3A and 16S rDNA primer pair HE1-HE3 did not produce visible products from individual ticks.

**Sequence variation among VLPT amplicons.** The amplicons were sequenced to confirm their identities and to characterize the sequence variations. Variations occurred at several levels, including the number and sequences of repeat units, the occurrence of a codon deletion (aspartic acid residue) immediately preceding the first repeat unit, the occurrence of a 9-bp deletion in the region downstream of the coding sequence, and single substitutions in certain nucleotide positions (Table 1 and Fig. 1).

Amplicons from different samples contained three to six repeat units, representing the four different sizes. To facilitate characterization of sequence differences, the deduced amino acid sequences of the repeat units from all amplicons were aligned, and type numbers were assigned for conserved sequences. There were five conserved types that differed at 4 to 18 amino acid positions. Two additional types (types 6 and 7) were each found in only one amplicon. The nucleotide sequences of individual repeat unit types were very conserved, with little variation in the third positions of codons. The deduced amino acid sequence of *E. chaffeensis* Wakulla, the only example of a six-repeat unit VLPT in our series, is shown in Fig. 3, and the repeat units are numbered to illustrate the five conserved types. A total of six different repeat profiles (number and order of repeat unit types) were represented among the

TABLE 1. Summary of VLPT sequence variation

Sample	Geographic origin	No. of repeat units	Repeat profile	A or G at position -69	A or G at position 6	A or G at position 27	A or G at position 487	Aspartic acid deletion	Presence of nine-base gap	No. of repeats in 120-kDa antigen gene <sup>a</sup>
Arkansas isolate	Arkansas	4	1, 2, 3, 4	A	A	G	G	No	No	4
91HE17 isolate	Arkansas	5	1, 2, 3, 3, 4	A	A	G	G	No	No	4
Sapulpa isolate	Oklahoma	3	1, 2, 4	A	A	G	G	No	No	3
Jax isolate	Florida	4	1, 2, 3, 4	A	A	G	G	No	No	4
St. Vincent isolate	Georgia <sup>b</sup>	3	1, 2, 4	A	A	G	G	No	No	3
Osceola isolate	Florida	4	1, 2, 3, 4	A	A	G	G	No	No	3
Wakulla isolate	Florida	6	1, 2, 3, 3, 4, 5	G	G	A	A	Yes	Yes	4
Liberty isolate	Florida	4	1, 2, 3, 4	G	G	G	G	No	No	4
West Paces isolate	Tennessee	3	1, 2, 4	A	A	G	G	No	No	3
Patient 1 blood	Arkansas	5	1, 2, 3, 4, 5	G	G	A	A	Yes	Yes	ND <sup>c</sup>
Patient 2 blood	Texas	5	1, 2, 3, 3, 4	A	A	G	G	No	No	ND
Patient 3 blood	Florida	5	1, 2, 3, 4, 5	G	G	A	A	Yes	Yes	ND
Patient 4 blood	Delaware	5	1, 2, 6, 4, 5	G	G	A	A	Yes	Yes	ND
Patient 5 blood	Maryland	4	1, 2, 3, 4	A	A	G	G	No	No	ND
Patient 6 blood	Florida	4	1, 2, 3, 4	A	A	G	G	No	No	ND
Patient 7 blood	Maryland	5	1, 2, 3, 4, 5	G	G	A	A	Yes	Yes	ND
Tick 97-36	Florida	4	1, 2, 3, 4	ND	A	G	ND	No	ND	ND
Tick pool 9	Missouri	5	1, 2, 3, 4, 5	G	G	A	G	Yes	No	4
Tick pool 20	North Carolina	4	1, 2, 3, 4	G	A	G	A	No	No	3
Tick pool 25	Kentucky	5	1, 2, 3, 3, 4	A	A	G	G	No	No	4
Tick pool 85	North Carolina	4	1, 2, 3, 4	G	A	G	A	No	Yes	3
Tick pool 95	North Carolina	3	1, 2, 4	A	A	G	A	No	Yes	3
Tick pool 97	North Carolina	4	1, 2, 3, 4	A	A	G	G	No	No	3
Tick pool 130	North Carolina	4	1, 2, 3, 7	A	G	A	A	Yes	No	4

<sup>a</sup> The 120-kDa antigen gene PCR results for tick pools were taken from Yu et al. (30).<sup>b</sup> Florida patient with history of tick exposure in Georgia.<sup>c</sup> ND, not done.

VLPT sequences amplified from isolates and patient whole blood (Table 1). Repeat types are listed in the downstream-to-upstream orientation. Type 1 represents the partial repeat always found the farthest downstream (Fig. 3). Among the amplicon sequences containing three or four repeats, the types and linear order of repeat units were conserved (e.g., 1, 2, 4 and 1, 2, 3, 4). Amplicons containing more than four repeat units usually contained a duplication of the type 3 unit or the addition of a type 5 unit. Among the amplicon sequences containing five repeats, we found three variations in the types and order of repeat units (Table 1).

In seven of the amplicons (representing one isolate, four patient blood samples, and two tick pools), three nucleotides representing an aspartic acid codon were absent. When present,

this aspartic acid residue was the 17th residue from the amino terminus and immediately preceded the beginning of the first repeat unit.

A gap consisting of a nine-nucleotide deletion (5'-GTTTT ATAT) occurred in the same position in seven amplicons (representing one isolate, four patient samples, and two tick pools). The G in this sequence is located 45 nucleotides downstream of the putative stop codon, and the sequence is marked in boldface type in Fig. 1. A repetitive motif (GTTTT) that may facilitate recombination was found in this region. The nine-base gap and the aspartic acid deletions were found more often in the amplicons with a larger number of repeats, and these deletions were always associated with each other in amplicons

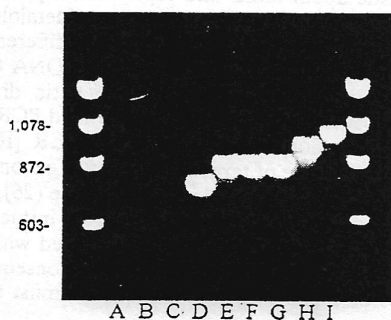


FIG. 2. Agarose gel electrophoresis of PCR products amplified from different cell culture isolates of *E. chaffeensis* and control samples with VLPT primers FB5A and FB3A. Lanes: A, water; B, extraction blank; C, uninfected DH82 cells; D, St. Vincent; E, Arkansas; F, Jax; G, Osceola; H, 91HE17; I, Wakulla. The outside lanes contain  $\phi$ X174/*Hae*III molecular mass markers.

TABLE 2. Collection information and PCR results for pools of adult *A. americanum* ticks

Tick pool no.	No. of ticks in pool	State of origin	16S rRNA PCR result <sup>a</sup>	VLPT PCR result	120-kDa gene PCR result <sup>b</sup>
9	10	Missouri	+	+	+
20	12	North Carolina	+	+	+
25	2	Kentucky	+	+	+
85	10	North Carolina	+	+	+
95	10	North Carolina	+	+	+
97	10	North Carolina	+	+	+
130	10	North Carolina	+	+	+
13	4	Missouri	-	-	ND <sup>c</sup>
21	2	Missouri	-	-	ND
219	8	Missouri	-	-	ND

<sup>a</sup> The 16S rRNA gene PCR results for positive pools were taken from Anderson et al. (4).<sup>b</sup> The 120-kDa antigen gene PCR results were taken from Yu et al. (30).<sup>c</sup> ND, not done.



\* MSQFSEDNMGNIQMPPD

(5) SNLQESSHLELPSLSEKVIHLESGLQQSSD

(4) SDSHEPSHLELPSLSEEVIQLES DLQSSN

(3) SDLHGFSFVELFDPFKEAVQLGNDLQSSD

(3) SDLHGFSFVELFDPFKEAVQLGNDLQSSD

(2) SDLHGPFVSVELFDPFKEEVQLES DLQSSN

(1) SDLHESSFVELPGPSKEEVQFEDDAKNVVP

GQDHVSLSELGLLLGGVFSTMNLYLSGYTPY

YYHHYCCNPPYYFDYVTPDYCHHCSESSLE

FIG. 3. Deduced amino acid sequence of the VLPT gene of *E. chaffeensis* Wakulla with the six repeat units aligned and numbered according to sequence type. Repeat units with conserved sequences were categorized and numbered to facilitate description of genetic variation. Repeat unit types and the order in which they occur vary among some strains.

derived from the isolates and patient blood samples but not the tick pools.

Adenosine or guanosine substitutions were found primarily at four distinct positions (positions -69, 6, 27, and 487) among the different VLPT amplicons (Table 1). Nucleotide substitutions at positions 27 and 487 resulted in amino acid substitutions from methionine to isoleucine and serine to glycine, respectively. Position 6 corresponds to the second nucleotide from the 3' end of primer FB5. We are currently evaluating the use of a primer slightly upstream of FB5, because sequence variation in the primer site could adversely affect PCR sensitivity.

In summary, among the three-repeat-unit varieties, the Sapulpa, St. Vincent, and West Paces isolates had identical sequences, but the sequence of the amplicon from tick pool 95 differed at position 487 and had the 9-base gap. Among the four-repeat-unit varieties, amplicons from the Arkansas, Jax, and Osceola isolates, patient 5, patient 6, and tick pool 97 had identical VLPT sequences, although the Osceola isolate had a three-repeat-unit version of the 120-kDa antigen gene and the others had the four-repeat-unit version. The sequences of the remaining four-repeat-unit versions, those from tick pools 20, 85, and 130 and the Liberty isolate, are different from those listed above and from each other. Among the five-repeat-unit varieties, amplicons from the 91HE17 isolate, patient 2, and tick pool 25 had identical VLPT sequences. Amplicons from patients 1, 3, and 7 had identical sequences, and the sequences from patient 4 and tick pool 9 were unique. The Wakulla isolate contained six repeat units.

**Expression of VLPT protein from cloned amplicons.** Amplicons from the St. Vincent, Jax, and 91HE17 isolates (representing the three-, four-, and five-repeat-unit versions, respectively) were cloned into an expression vector, and *E. coli* lysates were examined for expression of the VLPT protein. The VLPT genes of the Arkansas and Jax isolates were equal in size (four repeat units) and expressed comigrating proteins. The proteins expressed by the St. Vincent and 91HE17 clones were proportional to the sizes of the ORFs (Fig. 4). Expression of the VLPT was cytotoxic to *E. coli*, resulting in colonies approximately a quarter the size of those produced by the same strain of *E. coli* not transformed with a recombinant plasmid containing the VLPT insert.

**Amplification of 120-kDa antigen gene.** PCR amplification of the *E. chaffeensis* 120-kDa antigen gene was performed to

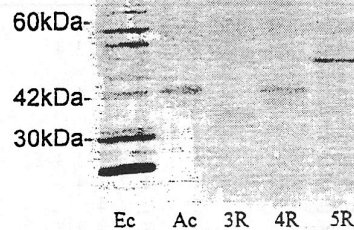


FIG. 4. Western immunoblot reacted with anti-*E. chaffeensis* HMAF showing the proteins expressed by three different size variations of the VLPT gene. The lanes contained lysates from purified *E. chaffeensis* Arkansas (lane Ec) *E. coli* transformed with a VLPT clone selected from the *E. chaffeensis* Arkansas genomic library (lane Ac), and *E. coli* transformed with plasmids containing PCR amplicons from the following strains: St. Vincent (lane 3R), Jax (lane 4R), and 91HE17 (lane 5R) (R indicates the number of 90-bp repeat units found in the VLPT gene of each strain).

additionally characterize each isolate, because variation among strains was described previously (30). Two size variants were found (Table 1), and these corresponded to the three- and four-repeat-unit versions described previously (6, 30) in the Sapulpa and Arkansas strains of *E. chaffeensis*, respectively. The Osceola, St. Vincent, and West Paces isolates produced amplicons of the same size as that produced by the Sapulpa strain, and the Jax, Wakulla, Liberty, and 91HE17 isolates produced amplicons of the same size as that produced by the Arkansas strain. These amplicons were partially sequenced with a single primer to confirm their identities. The corresponding patient blood samples were not tested. DNA samples from the tick pools were previously tested by PCR with the 120-kDa antigen gene primers (30), and one of the two size variations (e.g., three or four repeat units) was found in each pool (Table 1).

## DISCUSSION

In this report we present the results of the cloning, sequencing, and expression of a previously uncharacterized *E. chaffeensis* gene and describe a wide range of sequence variation among PCR amplicons derived from different isolates, patient samples, and ticks. The VLPT gene appears to be a sensitive and specific marker for *E. chaffeensis*. This gene has multiple diagnostic applications and represents a potentially important tool for the study of the molecular epidemiology of *E. chaffeensis*. The VLPT gene conveniently differentiates strains of *E. chaffeensis*. Some of the established DNA fingerprinting techniques for the detection of genetic diversity among strains of bacteria, such as arbitrary primed PCR (randomly amplified polymorphic DNA [RAPD] PCR [RAPD-PCR]), have the advantage of targeting the entire genome and require little knowledge of the nucleotide sequence (26). However, RAPD-PCR is difficult to apply to obligate intracellular bacteria, primarily because of problems associated with contamination of preparations with host-cell DNA. Consequently, conserved genes or regions of known sequence must be targeted.

Two of the first *Ehrlichia* genes sequenced, the 16S rRNA gene (1) and the *groESL* operon (24), appear to be completely conserved among several isolates of *E. chaffeensis* (19). Analysis of tRNA interrepeat length polymorphisms has been used

to detect and differentiate *Ehrlichia* species (9), but the utility of this method for the detection of differences among strains of the same species has not been determined. Two *E. chaffeensis* genes for which isolate-dependent sequence polymorphisms have been demonstrated are the 120-kDa antigen gene (29) and the VLPT described in this report. PCR amplification and nucleotide sequencing of the VLPT repeat region are convenient and more informative than PCR amplification and nucleotide sequencing of the 120-kDa antigen gene because of the smaller size of the VLPT repeat region and the greater variability in the number of repeat units. Finer resolution can be obtained by combining data obtained from assays with these genes. Interestingly, among the samples for which the number of repeats have been determined for both genes, there appears to be some correlation between gene sizes. In our study, the three-repeat-unit version of the 120-kDa antigen was always associated with fewer (e.g., three or four) repeat units in the VLPT. We have not detected any significant sequence homology between these two genes. The recently described p28 antigen gene family of *E. chaffeensis* may also prove to be useful for the study of strain variation (18), since variation was detected in a related gene family, *map-1* of *C. ruminantium* (21).

On a practical level, VLPT PCR is useful when conducting experiments involving different strains of *E. chaffeensis*. We have found it to be convenient to use DNA extracted from the Wakulla isolate as a positive PCR control. To date, the six-repeat version found in the Wakulla isolate appears to be relatively rare, and its use allows us to discern false-positive PCR results that might result from contamination of test samples with amplicons derived from a positive control. This is particularly important when nested PCR is used. Isolates and patient samples are easily identified by the VLPT size differences determined by gel electrophoresis of PCR amplicons. However, if sequencing is available, other variations at the nucleotide level may be discerned, allowing even finer separation. These variations include the presence or absence of the aspartic acid codon preceding the repeat region, the presence or absence of a 9-bp gap beyond the coding sequence, and individual nucleotide substitutions that occur at defined positions. Isolates, patient blood samples, and tick samples (24 total samples) could be segregated into 12 different groups by using information derived from individual VLPT sequences.

On an epidemiologic level, it would be interesting to find a correlation between the geographic origin of samples and variations in gene sequences. A previous report on the geographic distribution of variants of the 120-kDa antigen gene failed to demonstrate convincing geographic clustering of three- and four-repeat-unit variants among DNA samples from tick pools. All tick pools showed a single PCR band, suggesting that a single-repeat variant was present in each pool (30). We were not able to identify obvious patterns among VLPT sequences from similar geographic locations, but relatively few samples have been tested. The ticks included in each individual pool were collected in the same county, and we were interested to see whether we would find more than one version of the VLPT in a single pool. A single tick pool produced two amplicons that comigrated with the VLPT markers, suggesting the presence of two versions in the same pool; however, the second (less intense) band could not be recovered in sufficient quantity to verify its identity by sequencing. We are currently developing labeled probes for confirmation of amplicon identity by hybridization methods. Amplicons were separated by gel electrophoresis, but different sequence types with the same number of repeats would not have separated. Although the electropherograms did not reveal ambiguities, sequence data from the tick pools should probably be interpreted with more caution than

sequence data derived from isolates and patient blood samples. The VLPT PCR appears to be as sensitive as 16S rDNA PCR for the detection of *E. chaffeensis* and produces less background, particularly with ticks.

Current data do not allow us to determine the frequency with which VLPT sequences change or whether changes occur randomly or may be associated with the function of the protein. The wide geographic distribution of different sequence types may indicate frequent change. However, in a limited number of passages, isolate-specific sequence patterns remained stable in cell culture, suggesting that this gene could be used as a reliable marker to distinguish among isolates of *E. chaffeensis*.

The VLPT ORFs of the St. Vincent, Jax, and 91HE17 isolates code for polypeptides of 168, 198, and 228 amino acids, respectively. The corresponding calculated molecular masses (without posttranslational modification) are 19, 22, and 25 kDa. The apparent molecular masses of the VLPT proteins (Fig. 4) derived by electrophoretic mobility were approximately twice those calculated from the length of the ORF. The cause of this discrepancy remains to be determined. The *E. chaffeensis* 120-kDa antigen also migrates as though it has a molecular mass much higher than that predicted from the length of the ORF (29).

In our laboratory, characterization of the VLPT gene was concurrent with efforts to obtain new isolates of *E. chaffeensis*. The latter efforts produced six new isolates, and we naturally concentrated our efforts on using the VLPT gene as a tool for the differentiation of these isolates. Fundamental immunologic investigations on the role of this gene, the cellular location of the gene product, and its potential use as a diagnostic reagent remain.

#### ACKNOWLEDGMENTS

We thank Davis Janowski (Florida Department of Health) for guidance and assistance in collecting ticks; Scott Folks (Tallahassee Community Hospital), Lisa Rotz (CDC), and G. Merrill Shore (St. Vincent's Medical Center, Jacksonville, Fla.) for identifying the human monocytic ehrlichiosis patients from whom isolates were obtained; Pablo Manzowitz (National University of Tucuman, Argentina) for construction of VLPT subclones; Tom Ksiazek (CDC) for providing anti-*E. chaffeensis* hyperimmune mouse ascitic fluid David Walker (UTMB) for providing the Sapulpa and 91HE17 isolates; Burt Anderson (University of South Florida) for providing DNA extracted from *B. henselae* and *R. rickettsii*; Steven Stockham (University of Missouri-Columbia) for providing canine blood infected with *E. ewingii*; Robin Gager and Michael Levy (North Carolina State University) for providing *E. canis*; Will Goff (Animal Disease Research Unit, Agricultural Research Service, U.S. Department of Agriculture, Pullman, Wash.) for providing *A. marginale* DNA; Cynthia Holland (Antech Diagnostic Laboratories, Irvine, Calif.) for providing *E. risticii*; Yasuko Rikihisa (Ohio State University) for providing *E. muris*; Michael Zyzak (U.S. Naval Medical Research Detachment, Lima, Peru) for providing the *A. americanum* tick infected with *E. chaffeensis* by capillary feeding; and the staff of CDC's Biotechnology Core Laboratory for providing oligonucleotide primers.

#### REFERENCES

- Anderson, B., and G. McDonald. 1993. Construction of DNA libraries of A-T rich organisms using EcoRI star activity. *Anal. Biochem.* 211:325-327.
- Anderson, B. E., J. E. Dawson, D. C. Jones, and K. H. Wilson. 1991. *Ehrlichia chaffeensis*, a new species associated with human ehrlichiosis. *J. Clin. Microbiol.* 29:2838-2842.
- Anderson, B. E., J. W. Sumner, J. E. Dawson, T. Tzianabos, C. R. Greene, J. G. Olson, D. B. Fishbein, M. Olsen-Rasmussen, B. P. Holloway, E. H. George, and A. F. Azad. 1992. Detection of the etiologic agent of human ehrlichiosis by polymerase chain reaction. *J. Clin. Microbiol.* 30:775-780.
- Anderson, B. E., K. G. Sims, J. G. Olson, J. E. Childs, J. F. Piesman, C. M. Happ, G. O. Maupin, and B. J. Johnson. 1993. *Amblyomma americanum*: a potential vector of human ehrlichiosis. *Am. J. Trop. Med. Hyg.* 49:239-244.
- Centers for Disease Control and Prevention. 1998. Unpublished data.



6. Chen, S. M., X. J. Yu, V. L. Popov, E. L. Westerman, F. G. Hamilton, and D. H. Walker. 1997. Genetic and antigenic diversity of *Ehrlichia chaffeensis*: comparative analysis of a novel human strain from Oklahoma and previously isolated strains. *J. Infect. Dis.* 175:856-863.
7. Dame, J. B., S. M. Mahan, and C. A. Yowell. 1992. Phylogenetic relationship of *Cowdria ruminantium*, agent of heartwater, to *Anaplasma marginale* and other members of the order *Rickettsiales* determined on the basis of 16S rRNA gene sequence. *Int. J. Syst. Bacteriol.* 42:270-274.
8. Dawson, J. E., B. E. Anderson, D. B. Fishbein, J. L. Sanchez, C. S. Goldsmith, K. H. Wilson, and C. W. Dumtley. 1991. Isolation and characterization of an *Ehrlichia* sp. from a patient with human ehrlichiosis. *J. Clin. Microbiol.* 29:2741-2745.
9. Dawson, J. E., C. K. Warner, S. A. Ewing, S. R. Telford, R. E. Corstvet, R. Brennan, and J. G. Olson. 1997. Fingerprinting of *Ehrlichia* species by repetitive element polymerase chain reaction. *Am. J. Trop. Med. Hyg.* 57: 109-114.
10. Dumler, J. S., S. M. Chen, K. Asanovich, E. Trigiani, V. L. Popov, and D. H. Walker. 1995. Isolation and characterization of a new strain of *Ehrlichia chaffeensis* from a patient with nearly fatal monocytic ehrlichiosis. *J. Clin. Microbiol.* 33:1704-1711.
11. Eng, T. R., J. R. Harkess, D. B. Fishbein, J. E. Dawson, C. N. Greene, M. A. Redus, and F. T. Satalowich. 1988. Epidemiologic, clinical and laboratory findings of human ehrlichiosis in the United States. *JAMA* 264:2251-2258.
12. Ewing, S. A., J. E. Dawson, A. A. Kocan, R. W. Barker, C. K. Warner, R. J. Panciera, J. C. Fox, K. M. Kocan, and E. F. Blouin. 1995. Experimental transmission of *Ehrlichia chaffeensis* (Rickettsiales: Ehrlichieae) among white-tailed deer by *Amblyomma americanum* (Acari: Ixodidae). *J. Med. Entomol.* 32:368-374.
13. Fishbein, D. B., A. Kemp, J. E. Dawson, N. R. Greene, M. A. Redus, and D. H. Fields. 1989. Human ehrlichiosis: prospective active surveillance in febrile hospitalized patients. *J. Infect. Dis.* 160:803-809.
14. Gamble, W. C., W. A. Chappell, and E. H. George. 1978. Comparison of viral antibody titers of acid-precipitated and non-precipitated mouse ascitic fluid. *Health Lab. Sci.* 15:91-94.
15. Johnson, B., C. Happ, L. Mayer, and J. Piesman. 1992. Detection of *Borrelia burgdorferi* in ticks by species-specific amplification of the flagellin gene. *Am. J. Trop. Med. Hyg.* 47:730-741.
16. Lockhart, J. M., W. R. Davidson, D. E. Stallknecht, and J. E. Dawson. 1996. Site-specific geographic association between *Amblyomma americanum* (Acari: Ixodidae) infestations and *Ehrlichia chaffeensis*-reactive (Rickettsiales: Ehrlichieae) antibodies in white-tailed deer. *J. Med. Entomol.* 33:153-158.
17. Maeda, K., N. Markowitz, R. C. Hawley, M. Ristic, D. Cox, and J. E. McDade. 1987. Human infection with *Ehrlichia canis*, a leukocytic rickettsia. *N. Engl. J. Med.* 316:853-856.
18. Ohashi, N., N. Zhi, Y. Zhang, and Y. Rikihisa. 1998. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic gene family. *Infect. Immun.* 66:132-139.
19. Paddock, C. D., J. W. Sumner, G. M. Shore, D. C. Bartley, R. C. Elie, J. G. McQuade, C. R. Martin, C. S. Goldsmith, and J. E. Childs. 1997. Isolation and characterization of *Ehrlichia chaffeensis* strains from patients with fatal ehrlichiosis. *J. Clin. Microbiol.* 35:2496-2502.
20. Rechav, Y., M. Zyzak, L. J. Fielden, and J. E. Childs. Comparison of methods for introducing and producing artificial infection of ixodid ticks with *Ehrlichia chaffeensis*. *J. Med. Entomol.*, in press.
- 20a. Reddy, G. R. Personal communication.
21. Reddy, G. R., C. R. Sulsona, R. H. Harrison, S. M. Mahan, M. J. Burrridge, and A. F. Barbet. 1996. Sequence heterogeneity of the major antigenic protein 1 genes from *Cowdria ruminantium* isolates from different geographical areas. *Clin. Diagn. Lab. Immunol.* 3:417-422.
22. Reddy, G. R., C. R. Sulsona, A. F. Barbet, M. M. Suman, M. J. Burrridge, and A. R. Alleman. 1998. Molecular characterization of a 28 kDa surface antigen gene family of the tribe Ehrlichieae. *Biochem. Biophys. Res. Commun.* 247:636-643.
23. Staden, R. 1994. The STADEN package. *Methods Mol. Biol.* 25:9-170.
24. Sumner, J. W., K. G. Sims, D. C. Jones, and B. E. Anderson. 1993. *Ehrlichia chaffeensis* expresses an immunoreactive protein homologous to the *Escherichia coli* GroEL protein. *Infect. Immun.* 61:3536-3539.
25. Sumner, J. W., W. L. Nicholson, and R. F. Massung. 1997. PCR amplification and comparison of nucleotide sequences from the *groESL* heat shock operon of *Ehrlichia* species. *J. Clin. Microbiol.* 35:2087-2092.
26. Vaneechoutte, M., and J. Van Eldere. 1997. The possibilities and limitations of nucleic acid amplification technology in diagnostic microbiology. *J. Med. Microbiol.* 46:188-194.
27. van Vliet, A. H. M., F. Jongejan, M. van Kleef, and B. A. M. van der Zeijst. 1994. Molecular cloning, sequence analysis, and expression of the gene encoding the immunodominant 32-kilodalton protein of *Cowdria ruminantium*. *Infect. Immun.* 62:1451-1456.
28. Walker, D. H., and J. S. Dumler. 1996. Emergence of the ehrlichioses as human health problems. *Emerg. Infect. Dis.* 2:18-29.
29. Yu, X. J., P. Crocquet-Valdes, and D. H. Walker. 1997. Cloning and sequencing of a gene for the 120-kDa immunodominant protein of *Ehrlichia chaffeensis*. *Gene* 184:149-154.
30. Yu, X., J. F. Piesman, J. G. Olson, and D. H. Walker. 1997. Geographic distribution of different genetic types of *Ehrlichia chaffeensis*. *Am. J. Trop. Med. Hyg.* 56:679-680.