

Molecular Heterogeneity of *Ehrlichia chaffeensis* Isolates Determined by Sequence Analysis of the 28-Kilodalton Outer Membrane Protein Genes and Other Regions of the Genome†

Chuanmin Cheng,¹ Christopher D. Paddock,² and Roman Reddy Ganta^{1*}

Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, Manhattan, Kansas 66506,¹ and Centers for Disease Control and Prevention, Atlanta, Georgia 30333²

Received 4 September 2002/Returned for modification 8 October 2002/Accepted 16 October 2002

Ehrlichia chaffeensis, a tick-transmitted rickettsial agent, is responsible for human monocytic ehrlichiosis (HME). In this study, we genetically mapped 10 isolates obtained from HME patients. Sequence analysis of the 28-kDa outer membrane protein (OMP) multigene locus spanning 6 of the 22 tandemly arranged genes identified three distinct genetic groups with shared homology among isolates within each group. Isolates in Groups I and III contained six genes each, while Group II isolates had a gene deletion. There were two regions on the locus where novel gene deletion or insertion mutations occurred, resulting in the net loss of one gene in Group II isolates. Numerous nucleotide differences among genes in isolates of each group also were detected. The shared homology among isolates in each group for the 28-kDa OMP locus suggests the derivation of clonal lineages. Transcription and translation analysis of the locus revealed differences in the expressed genes of different group isolates. Analysis of the 120-kDa OMP gene and variable-length PCR target gene showed size variations resulting from loss or gain of long, direct repeats within the protein coding sequences. To our knowledge this is the first study that looked at several regions of the genome simultaneously, and we provide the first evidence of heterogeneity resulting from gene deletion and insertion mutations in the *E. chaffeensis* genome. Diversity in different genomic regions could be the result of a selection process or of independently evolved genes.

Human monocytic ehrlichiosis (HME), first reported in 1987, is caused by *Ehrlichia chaffeensis* (4, 10, 16). Early symptoms of HME include fever, headache, malaise, muscle aches, vomiting, diarrhea, cough, joint pains, and confusion (33). Untreated ehrlichiosis can become severe and is potentially fatal in immune-compromised and elderly people (22, 23). *E. chaffeensis* also infects other vertebrate hosts, including dogs, goats, and white-tailed deer (2, 5, 6, 8, 14), and shares close genetic similarity with *Ehrlichia canis*, *Ehrlichia ruminantium*, and *Ehrlichia ewingii* (9).

A multigene locus that encodes a 28-kDa outer membrane protein(s) (OMP) has been reported from *E. chaffeensis*, *E. canis*, and *E. ruminantium* (15, 18–21, 24–26, 34). The protein-coding sequences of the genes have four long stretches of conserved regions separated by three highly variable regions (VRs) where the dominant immunogenic B-cell epitopes are located (25). This locus has generated considerable interest for its possible role in immune evasion because the 28-kDa OMP locus shares structural similarity to antigenic variant surface antigen genes of *Borrelia burgdorferi* and *Neisseria gonorrhoeae* (12, 28, 36).

The protein-coding region of the gene encoding a 120-kDa OMP contains two to four nearly identical, highly hydrophilic 80-amino-acid tandem repeats (30, 35). The number of repeats

varies among different isolates, resulting in the size variations in the encoded protein. Similarly, within the coding region of the variable-length PCR target (VLPT) gene there is a variable number of direct nucleotide repeats that may code for various numbers of 30-amino-acid repeats (23, 31). The presence of variable direct repeats in *E. chaffeensis* is similar to that of the major antigenic variant surface protein of *Mycoplasma hominis* (37). *M. hominis* surface protein, termed variable adherence-associated antigen, contains one to four nearly identical repeats of 121 amino acids, and the gain or loss of repeats gives rise to distinct antigenic variants with size variations in variable adherence-associated antigen in clonal populations (37).

In this study we mapped *E. chaffeensis* isolates to examine variability in the genome. Specifically, the 28-kDa gene locus spanning 53 kb of DNA from 10 human isolates was characterized at the molecular level. We also compared the sequence data generated from 15 kb of the 120-kDa OMP gene and 4 kb of the VLPT gene from all 10 isolates.

MATERIALS AND METHODS

In vitro cultivation of *E. chaffeensis* isolates. Ten *E. chaffeensis* isolates obtained from whole blood or bone marrow of acutely ill patients with HME (Table 1) were cultivated in the canine macrophage cell line DH82, as described previously (3). All isolates were obtained from moderately to seriously ill patients, including two patients who died from the infection. Three isolates, Lithonia, Chattanooga, and Heartland, are new isolates reported in this study. The remaining seven isolates were reported previously (4, 23, 31). Cultured bacteria were harvested when 80 to 100% of the confluent DH82 cells were infected (11).

DNA filter hybridization analysis. Genomic DNA from all 10 isolates was purified by the sodium dodecyl sulfate (SDS) proteinase K-phenol-chloroform extraction method (17). The genomic DNA samples were digested with *EcoRI*, *EcoRV*, *HindIII*, or *XbaI* and were resolved on a 0.9% agarose gel and trans-

* Corresponding author. Mailing address: Department of Diagnostic Medicine/Pathobiology, College of Veterinary Medicine, Kansas State University, 1800 Denison Ave., Manhattan, KS 66506. Phone: (785) 532-4612. Fax: (785) 532-4851. E-mail: rganta@vet.ksu.edu.

† This paper is published as Kansas Agricultural Experiment Station Contribution number 02-497-5.

TABLE 1. *E. chaffeensis* isolates

Isolate name	Isolate origin	Date established	Reference	Genetic group ^a
Osceola	Florida	1997	31	I
Arkansas	Arkansas	1990	4	I
Lithonia	Georgia	2000	This paper	I
Chattanooga	Tennessee	2000	This paper	II
West Paces	Tennessee	1998	31	II
Heartland	Nebraska	1999	This paper	II
St. Vincent	Georgia	1996	23	II
Wakulla	Florida	1997	31	II
Liberty	Florida	1998	31	III
Jax	Florida	1996	23	III

^a Genetic grouping established in this study.

ferred onto a Hybond-N⁺ nylon membrane (Amersham, Piscataway, N.J.) by the capillary transfer method (17). The membrane was hybridized alternatively with ³²P-labeled gene probes for the 28-kDa OMP gene, followed by the 120-kDa OMP gene, the VLPT gene, and the 16S rRNA gene, respectively. A ~7-kb fragment spanning genes 14 through 19 of the Arkansas isolate generated by PCR (described in the next paragraph) was used as the probe for mapping the 28-kDa locus. Similarly, amplicons generated from the Arkansas isolate for 120-kDa and VLPT genes by using primers described in the next paragraph were used to make hybridization probes. A 0.39-kb rRNA gene segment amplified from the Arkansas isolate as described previously (11) was used to map the rRNA gene locus. Hybridization was performed overnight at 68°C. The blots were washed once each for 30 min at 68°C with 6× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate) containing 0.1% SDS, 2× SSC containing 0.1% SDS, 1× SSC containing 0.1% SDS, and 0.2× SSC containing 0.1% SDS, respectively. After a final stringent wash, membranes were exposed for 2 days to X-ray film at -70°C with an intensifying screen, and the film was developed in a Kodak film processor.

Gene analysis by PCR and nucleotide sequence. A 28-kDa OMP locus-specific primer pair was designed on the basis of the published sequence for the Arkansas isolate (RRG14, 5'CCGTTTCTGCTTTATTAGAATG; and RRG18, 5'RNCTAATAATTACAATGTGTG). The primer pair was used to amplify ~7-kb fragments spanning the genes 14 through 19 (34). PCR was performed by using a long PCR reagent kit (Gibco BRL, Rockville, Md.). High-fidelity DNA polymerase included in the PCR kit minimizes errors in amplified products. The PCR cycles, performed in a GenAmp9700 (Applied Biosystems, Foster City, Calif.), included one initial cycle at 94°C for 4 min followed by 35 cycles of 94°C for 30 s, 55°C for 30 s, and 68°C for 7 min and one extension cycle at 72°C for 15 min. Ten percent PCR products each were resolved on a 0.9% agarose gel and were detected by ethidium bromide staining. The primer pair for amplifying the 120-kDa gene was F1 (5'GAGAATTGATTGTGGAGTTGG), reported by Yu et al. (35), and RRG24 (5'CTATCTCAAGACTAAACCTTAC). RRG24 is a

downstream, reverse primer designed from the sequence reported by Yu et al. (35). VLPT gene-specific PCR primers FB5 (5'AAATAGGGTATAAATATGTCAC) and FB3 (5'GCCTAATTCAGATAAACTAAC) were used as described earlier (31).

PCR products of 28-kDa, 120-kDa OMP, and VLPT gene segments were purified by use of a PCR product presequencing kit (USB Corp., Cleveland, Ohio) and were used for sequence determination by primer walking with the manual Thermo Sequenase Radiolabeled Terminator Cycle Sequencing kit (USB Corp.). Sequence analysis was performed by using the GCG program package (7) available on the Kansas State University Unix computer system. The sequence data were deposited in the GenBank database.

Detection of RNA from the 28-kDa OMP gene locus by RT-PCR. Gene-specific primers, designed from the VRs within each gene for genes 14 to 19, were used in reverse transcription (RT)-PCR analysis (Table 2). Total RNA was extracted from *E. chaffeensis* cultures by use of the RNAwiz total RNA isolation kit (Ambion Inc., Austin, Tex.). RNA samples were stored at -70°C until use. Total RNA was treated with RQ1 DNase (Promega Corp., Madison, Wis.) to eliminate genomic DNA prior to use in RT-PCR assays. To increase the activity, the DNase treatment was performed for 1 h at 37°C in buffer provided by the vendor. In addition, 1 mM CaCl₂ and 1.5 mM MgSO₄ were added. The presence of gene-specific RT-PCR products was verified after transferring the products to a nylon membrane followed by hybridization with gene-specific probes.

Western blot analysis. Antigens used in Western blot analysis included whole-cell antigens from the *E. chaffeensis* Arkansas isolate and purified recombinant proteins for 28-kDa OMP genes 16 and 19 (formerly known as ORF2 and ORF5, respectively) of *E. chaffeensis* Arkansas isolate and an *E. canis* homologous (ORF1) (25). The recombinant proteins were prepared by using a prokaryotic expression system in *Escherichia coli* as previously described (24). Hyperimmune sera from B6 mice obtained after 50 days postinfection with *E. chaffeensis* (Arkansas isolate) (11) were used as the antibody source. The Western blot experiment was performed by using diluted (1:128) mouse serum as the primary antibody (11).

Nucleotide sequence accession numbers. Sequences reported in this paper were deposited in the GenBank database under numbers AF479833 to AF479840, AF474890 to AF474899, AF470688 to AF470697, AY117396, and AY117397.

RESULTS

DNA filter hybridization analysis. Southern blot analysis of 10 *E. chaffeensis* isolates with a 28-kDa OMP gene probe showed extensive restriction enzyme site differences. Isolates with similar restriction enzyme site patterns were grouped, and restriction-digested DNAs were resolved by groups having similar restriction maps (Fig. 1A). These analyses revealed the presence of restriction site differences that can be grouped into

TABLE 2. Primers used for RT-PCR analysis of 28-kDa OMP genes

Primer name (forward/reverse primers)	Gene amplified	Sequence and location of:	
		Forward primer	Reverse primer
RRG29/RRG71	14 of Group I	5' GACGCAATAGCAGATAAG (VRII)	5' GAGCTCCTTCTAATACTAC (VRIII)
RRG36/RRG72	15 of Group I	5' TACTGTGCGGTTGTATGGTTG (VRI)	5' GCACGTAGTACTAGCTGTG (VRII)
RRG73/RRG38comp	16 of Group I	5' GTGTAATATCTAGAACCACCTTAAGC (VRII)	5' CAGACGCAGCTGCCTGCACCATC (VRIII)
RRG74/RRG39comp	17 of Group I	5' CTCATCAAGTCACAATGATAATC (VRI)	5' GGTATTCGCTGTTGTCTTGTG (VRII)
RRG32/RRG76	18 of Group I	5' CACAATATCTAAAAATCTCCAG (VRI)	5' TAGCTTCCCCACTGTTATG (VRII)
RRG34/RRG77	19 of Group I	5' GAAGCGCAATATCCAACCTCCTC (VRI)	5' CTACTCATGTCTGCTGCTGAG (VRII)
RRG29/RRG71	14 of Group II	5' GACGCAATAGCAGATAAG (VRII)	5' GAGCTCCTTCTAATACTAC (VRIII)
RRG47/RRG83	15 of Group II	5' ATGCTACTGTGCTATTGTATGG (VRI)	5' CAGATGTGGCATTGCTACCAG (VRII)
RRG41/RRG82	16 of Group II	5' GATGGTGCAAATGATGCGTC (VRII)	5' GCAAGCGCTGATCCACTAG (VRIII)
RRG86/RRG59	17 of Group II	5' ACCAACGCTACACAAAAC (VRII)	5' AGGGTAGGAATATCTCG (VRIII)
RRG88/RRG90	19 of Group II	5' CTCCATTTACTGTTTCA (VRI)	5' TATTATCTGCTGCTATTGTG (VRII)
RRG29/RRG71	14 of Group III	5' GACGCAATAGCAGATAAG (VRII)	5' GAGCTCCTTCTAATACTAC (VRIII)
RRG48/RRG84	15a of Group III	5' TCGTACTGTGCGACTGTATGG (VRI)	5' TATTTGTTAGCTGTAGGATTGGTAC (VRII)
RRG93/RRG100	15 of Group III	5' GTACTAGTGCCACAGTAAATAAC (VRII)	5' AAGTCTGGATGCTGCAGATGATG (VRIII)
RRG42/RRG85	16 of Group III	5' GATGGTGCAAATGATGCGTC (VRII)	5' CTGCAAGCGCTGATTACTAG (VRIII)
RRG56/RRG87	17 of Group III	5' GTCAACAAGACAGCAGC (VRI)	5' GACGTAGCAAATGCTTTC (VRIII)
RRG89/RRG91	19 of Group III	5' GTACAACAGCTGGAGTATTG (VRI)	5' CTGAAGTGTGATGAGATAAAG (VRII)

three genetic groups, namely, Groups I, II, and III. Group I contained three isolates (Arkansas, Osceola, and Lithonia), Group II included five isolates (St. Vincent, Chattanooga, West Paces, Heartland, and Wakulla), and Group III included the Liberty and Jax isolates. The grouping is based on the recognition of nearly identical restriction site patterns for *EcoRI*- and *XbaI*-digested DNA. The 28-kDa gene probe in Group I isolates detected three *EcoRI* fragments, while Group II and III isolates had one and two restricted segments, respectively. Similarly, *XbaI*-restricted fragments were different in each group of isolates. However, the Wakulla isolate shared some overlap in its presence or absence of restriction sites with both Group II and Group III isolates for *EcoRV* and *HindIII* restriction sites. For example, in *EcoRV*-restricted DNA, a 0.65-kb fragment is common in all five isolates that were identified as Group II, whereas the 1.65-kb fragment is unique to St. Vincent, Chattanooga, West Paces, and Heartland isolates but not to the Wakulla isolate. Similarly, the Wakulla isolate contained the 9.4- and 5.0-kb *EcoRV* fragments similar to Group III isolates. It also had overlapping *HindIII* restriction sites as in other Group II isolates and Group III isolates (Fig. 1A). DNA blot analysis was also performed by using probes specific to three genomic regions spanning the 120-kDa OMP, VLPT, and 16S rRNA gene loci. Unlike the 28-kDa OMP locus, these three regions had fewer restriction site polymorphisms, and the differences did not result in clear grouping of isolates (Fig. 1B to D).

Sequence analysis of the 28-kDa OMP locus. The 28-kDa OMP multigene locus has been reported by three research groups, including ours, from *E. chaffeensis* and *E. canis* (15, 18–21, 24–26, 34). Evidence was presented recently for the presence of 22 tandemly arranged paralogous genes in this locus (15, 19, 34). Yu et al. and Long et al. (15, 34) named the genes in the locus spanning from the 5' end to 3' end p28-1' and p28-1 to p28-21. Ohashi et al. (19) gave some genes a designation beginning with OMP-1 followed by discontinuous letters, and they classified others from the locus as p28 followed by numbers. In previous publications genes from this locus were referred to as 28-kDa gene open reading frames (ORFs) (24, 25). In this study, we opted to follow the numbering system previously described (15, 34) because it is easier to follow. Genes 14 to 19 of the 28-kDa OMP locus described in this study are the same as those listed as genes 14 to 19 for the Arkansas isolate reported by Yu et al. (34). In the terminology of Ohashi et al. (19), they correspond to genes OMP-1B, OMP-1C, OMP-1D, OMP-1E, OMP-1F, and p28.

Approximately 7-kb-long DNA segments spanning six genes of the 28-kDa OMP locus (genes 14 to 19 of *E. chaffeensis* Arkansas isolate) were amplified from DNA of all 10 *E. chaffeensis* isolates (Fig. 2). Predicted-size PCR products were detected in Group I and III isolates, while the amplicons for the Group II isolates, including the Wakulla isolate, were about 1.2 kb smaller. The reduction in the size of Group II PCR fragments was approximately equal to the loss of one gene. The entire protein-coding sequence of the 120-kDa gene and partial segments of the VLPT gene were also amplified. Unlike the DNA blot analysis, amplification of these two gene segments showed notable size differences (Fig. 2).

Nucleotide sequence data for the 28-kDa gene amplicons are illustrated in Fig. 3 (complete sequence data were submit-

ted to GenBank). The tandemly arranged paralogous genes of each isolate have much greater variation than orthologous genes of different isolates. Most of the variation was located within the three VRs of the coding sequences. In addition, two regions of the locus where the gene deletion or insertion mutations occurred resulted in the net loss of one gene in Group II isolates (Fig. 3). Gene 18 was unique to Group I isolates, while gene 15a was identified only in Group III isolates. Sequence comparisons of the coding regions at the amino acid level among the orthologous genes from different genetic group isolates had homology that ranged from 82 to 97% (gene 14, 97%; gene 15, 82 to 93%; gene 16, 96%; gene 17, 88 to 97%; gene 19, 91 to 93%). This homology was higher than that observed among the tandemly arranged paralogous genes within each isolate (53 to 88%; Group I isolates, 55 to 87%; Group II, 53 to 84%; Group III, 54 to 88%). The most closely related paralogs were genes 15, 15a, and 17, as judged by phylogenetic analysis (Fig. 4).

Comparison of sequences among isolates from Groups I and III revealed only one difference each, located in VRIII of gene 15 (Group I) and VRI of gene 15a (Group III). The mutation in Group I isolates Arkansas to Osceola resulted in an amino acid change from Ser to Pro, while the mutation in Group III did not lead to amino acid changes. Significantly, more differences were observed among the Group II isolates St. Vincent, Chattanooga, West Paces, Heartland, and Wakulla. The isolates West Paces and Heartland had identical sequences. These two isolates and Wakulla, however, differ from St. Vincent at two positions in gene 14 that lead to one amino acid change from Phe to Val. In gene 16 they also differed at two nucleotide positions. The first mutation translated to an amino acid change of Tyr to Asn, while the second mutation occurred in the termination codon of St. Vincent (*taa* to *tta*) resulting in the addition of four new amino acids at the 3' end of West Paces, Heartland, and Wakulla isolates. St. Vincent also had one more mutation than West Paces and Heartland. It changed amino acid Ile to Val. In gene 17, St. Vincent differed from Wakulla at two positions, causing one amino acid change from Leu to Met. In gene 19, West Paces, Heartland, and St. Vincent isolates differed from Wakulla at four positions, causing four amino acid differences from Gln, Ser, Lys, and Asn to His, Asn, Thr, and Ser, respectively. The mutations included both transversions and transitions.

To verify the organization of the mapped region, specifically the position of each gene in the locus and the loss or gain of genes among isolates, PCR analysis of several overlapping fragments was performed (Fig. 5). PCR primers were designed from the variable regions within the protein-coding sequences of the 28-kDa gene locus (Table 2), and the sizes of the expected amplicons were estimated (Fig. 5A). The sizes of the amplicons generated by using these primer sets were evaluated (Fig. 5B). The amplicons generated by using these primer sets were of the same sizes as expected, thus verifying the position of each gene on the mapped locus.

Sequence analysis of the 120-kDa OMP and VLPT genes. PCR products for the 120-kDa and VLPT genes were also sequenced. The data were presented schematically in Fig. 6 (these data were also deposited in the GenBank database). Sequence analysis revealed differences primarily resulting from the loss or gain of 240-bp-long repeats for the 120-kDa gene

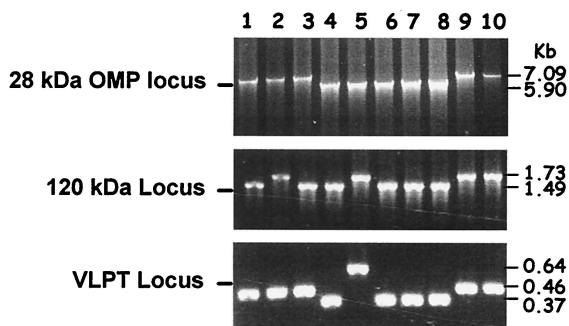


FIG. 2. PCR analysis of three gene loci of *E. chaffeensis*. (A) The 28-kDa OMP loci spanning genes 14 through 19 were amplified, and the products were resolved on a 0.9% agarose gel and detected after being stained with ethidium bromide. Lanes 1 to 10 contained PCR products generated from Osceola, Arkansas, Lithonia, Chattanooga, Wakulla, West Paces, Heartland, St. Vincent, Liberty, and Jax isolates, respectively. Panels B and C are the same as panel A except that the PCR primers used for amplification were specific to the 120-kDa gene and VLPT gene segments, respectively.

and 90-bp-long repeats for VLPT genes. The 120-kDa gene had three to four repeats, whereas the VLPT gene contained three to six repeats. The isolate grouping based on these two genes was different and did not correlate with the 28-kDa gene-based grouping of isolates. These differences were also apparent in the size variations observed in the amplicons of all three gene loci (Fig. 2) but not in the Southern blot data (Fig. 1), possibly due to large fragment size of the restriction-digested fragments.

Gene expression of the 28-kDa OMP gene locus. To determine the transcriptional activity of the 28-kDa gene locus, RT-PCR analysis was performed by using gene-specific primer pairs (Table 2). The analyses were carried out on several independently made preparations of RNA obtained from cultured organisms of one isolate each from Group I (Arkansas) and Group III (Liberty) and two isolates from Group II (St. Vincent and Wakulla). Typical data for the Arkansas isolate are presented in Fig. 7, and the summary of analyses performed on all four isolates are in Table 3. This experiment revealed differences in the genes expressed in different genetic groups. Differences also were noted for two isolates when analyzed by using different batches of RNA made from cultures harvested on different culture days. For the Group I isolate Arkansas, RT-PCR products were detected only for genes 14, 15, 18, and 19 but not for genes 16 and 17. All genes, except gene 16, contained transcripts for the Group II isolate St. Vincent. The expression pattern remained unchanged for these two isolates. Another Group II isolate, Wakulla, contained variable numbers of transcripts that ranged from two to five genes. Similarly, a Group III isolate, Liberty, had variations in the expressed transcripts that fluctuated from the expression of three to six genes. Specificity of the amplicons to

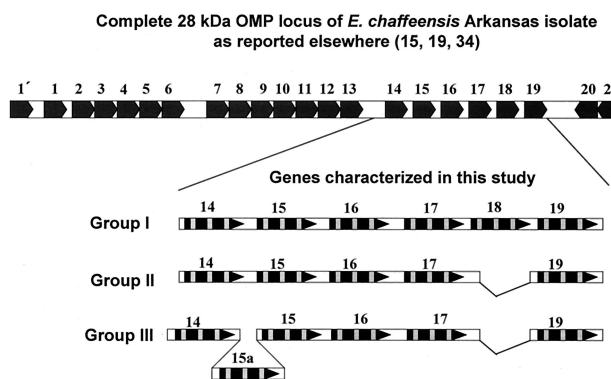


FIG. 3. Comparison of nucleotide sequences of the 28-kDa gene amplicons of all three genetic group isolates. Sequences for genes 14 to 19 of the amplified 28-kDa OMP locus segments were generated by Sanger's manual dideoxy chain termination method with synthetic overlapping primers. The entire gene locus spanning all 22 genes present in the genome for the Arkansas isolate is presented at the top of the figure for comparison. Arrowheads indicate the orientation of the protein coding sequence. Protein coding sequences are identified with black boxes, and variable regions within the coding regions are highlighted by gray boxes.

RNA was confirmed by including a no-reverse-transcriptase control in every experiment (Fig. 7). Lack of RT-PCR products for some genes was not the result of nonpriming of primers, because they successfully amplified the predicted-size fragments when using genomic DNA as templates (Fig. 7). Similarly, it was not the result of poor quality RNA, because all batches of RNA used in this study successfully served as templates in an RT-PCR assay for the rRNA gene (data not shown).

Western blot analysis was performed to examine protein expression from the 28-kDa OMP locus. This experiment was performed by using the Arkansas isolate, because RNA expression of the genes analyzed remained constant (described above). In this assay, purified recombinant proteins for a non-transcribed gene (gene 16) and a transcribed gene (gene 19) and a whole-cell protein extract prepared from the cultured organisms were used as the antigens. An *E. canis* homologue of the 28-kDa OMP gene recombinant protein was also included as a control. Serum obtained from a B6 mouse after infection with *E. chaffeensis* Arkansas isolate was the antibody source. The Western blot data in Fig. 8 shows the presence of reactive antibodies for the gene 19 antigen and antigens prepared from a whole-cell lysate of *E. chaffeensis* but not for the gene 16 antigen and a homologous gene from *E. canis* (Fig. 8).

DISCUSSION

In the 10 *E. chaffeensis* isolates examined, we identified extensive restriction fragment length polymorphisms in the ge-

FIG. 1. DNA filter hybridization of *E. chaffeensis*. Ten isolates from HME patients were used: Osceola, lane 1; Arkansas, lane 2; Lithonia, lane 3; Chattanooga, lane 4; West Paces, lane 5; Heartland, lane 6; St. Vincent, lane 7; Wakulla, lane 8; Liberty, lane 9; Jax, lane 10. DNA from each isolate was digested to completion, resolved on a 0.9% agarose gel, transferred to a nylon membrane, and hybridized with ³²P-labeled probes specific to the 28-kDa gene locus (A); 120-kDa gene (B); VLPT gene (C); or ribosomal RNA gene (D). Restriction enzymes used for the DNA blot analysis are identified on the top of each blot. Molecular size markers are present on the left side of the profiles.

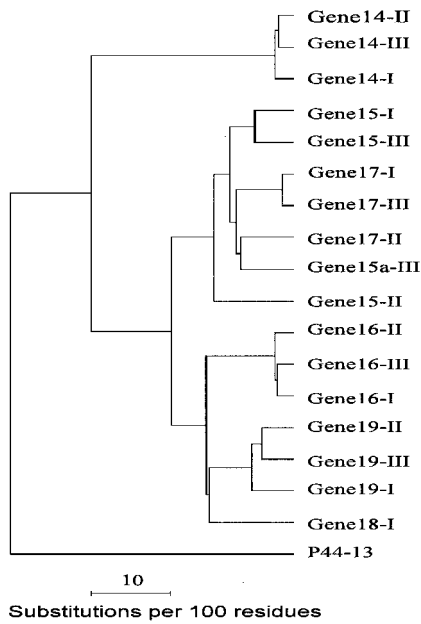


FIG. 4. Phylogenetic tree of the 28-kDa OMP locus spanning orthologous and paralogous genes 14 to 19. Protein-coding sequences that encode 28-kDa OMPs from all genes depicted in Fig. 3 were translated and used to construct the phylogenetic tree. Sequences from one isolate each (Arkansas, Jax, and St. Vincent) were used to prepare the tree with the GCG programs Pileup, Distance, and Growtree (7). An OMP, p44-13, a sequence of *Anaplasma phagocytophila* (GenBank accession no. AF414592), was used as an outgroup to root the tree. Orthologous genes 15, 15a, and 17 clustered on the same branch of the tree, while genes 14, 16, 18, and 19 grouped on separate branches. Symbols I, II, and III in the figure represent genes from isolates of Groups I, II, and III, respectively.

nome spanning the 28-kDa OMP multigene locus. On the basis of the shared genetic similarity, judged from restriction enzyme analysis together with sequence data of the 28-kDa OMP locus, we grouped the isolates into three distinct genetic groups, Groups I, II, and III. Little sequence variation was observed among different isolates within each group for the region of the 28-kDa locus. Most of the differences were present in Group II isolates. On the basis of *EcoRI*- and *XbaI*-restricted fragment analysis, the Wakulla isolate grouped with other Group II isolates while the data for *EcoRV* and *HindIII* digests also showed similarities with Group III isolates at some sites. The observed overlap of restriction pattern of the Wakulla isolate with both Group II and Group III agents may suggest that it is evolving from Group II. Since the sequence data of the Wakulla isolate had a greater homology with other Group II isolates, we identified it as a Group II isolate. Three distinct groups of *E. chaffeensis* isolates based on the sequence comparison of gene 19 orthologs also were reported recently (15). The isolates utilized for that analysis (15) included two isolates, Arkansas and Jax, also analyzed in our study. These observations independently verify the existence of different clonal populations of *E. chaffeensis* in nature. Taken together, these studies support that *E. chaffeensis* isolates are derived from at least three clonal lineages as judged by the analysis of the 28-kDa OMP locus, and the isolates within each group may represent different strains.

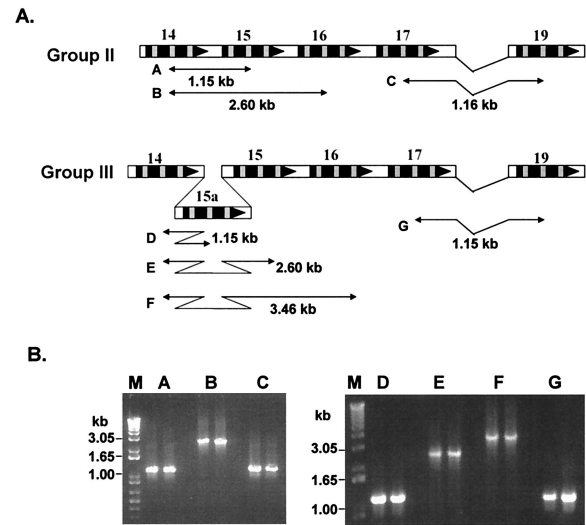


FIG. 5. Verification of the mapped 28-kDa locus from Group II and III isolates. To verify the organization of genes, PCR primers were designed to amplify predicted size overlapping fragments. (A) The primers were designed from the variable regions of coding sequences, and the sizes of the predicted amplicons were estimated. Segments A and B verify the location of genes 14 and 15 for Group II isolates. Segments C and G verify the loss of gene 18 in Group II and III isolates. The gain of gene 15a in Group III isolates was verified by amplifying segments D, E, and F. (B) Amplification data for two isolates each (Group II, Wakulla and St. Vincent; Group III, Liberty and Jax) are presented. The predicted amplicons were detected in the PCR products.

We and other researchers have reported that the protein coding sequence of the genes in the 28-kDa OMP locus are divided into three highly variable regions that are separated by four highly conserved regions (15, 18–21, 24–26, 34). The variable regions constitute hydrophilic regions and may represent B-cell epitopes (25). In this study, we observed a similar gene structure having conserved and variable regions in the protein-coding sequences of paralogous genes of all 10 isolates. Vari-

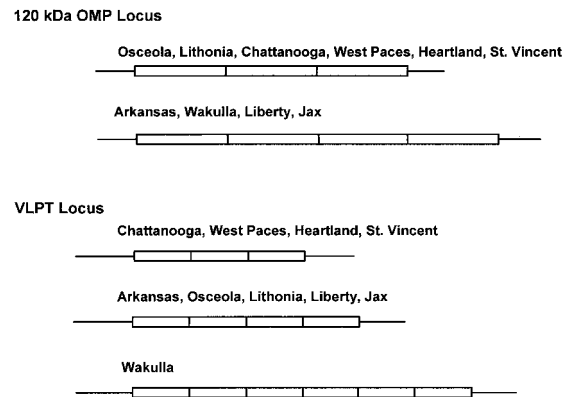


FIG. 6. Schematic representation of the nucleotide sequence data generated from the 120-kDa OMP and VLPT gene loci. Lines represent unique sequences at both 5' and 3' ends of the amplified segments. Boxes represent tandem repeats. The repeat sizes in 120-kDa and VLPT genes are 240 and 90 bp each, respectively. The number of repeats vary in different isolates which is represented by the varying numbers of boxes.

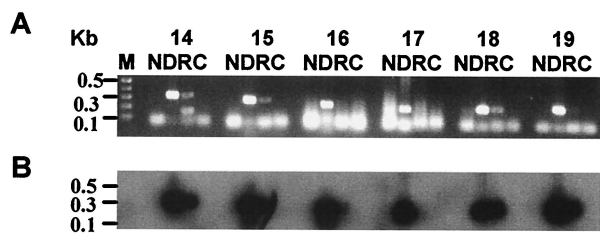


FIG. 7. RT-PCR analysis of *E. chaffeensis* isolate Arkansas RNA extracted from in vitro-cultured organisms. Purified RNA was digested with RQ1 DNase for 1 h to remove any residual genomic DNA. DNA-free RNA was used for the RT-PCR assay by using gene-specific primer sets for genes 14 to 19. The primers were designed from variable regions and are listed in Table 2. Lane N contained all reagents for the RT-PCR assay, except RNA. Lane D is the same as lane N plus genomic DNA. Lane R is same as lane N but with DNase-treated RNA as the template. Lane C is same as lane R except that the reverse transcriptase is omitted. (A) Amplified products resolved on a 1.2% agarose gel and detected after being stained with ethidium bromide. (B) RT-PCR products transferred to a nylon membrane and hybridized with a ³²P-labeled 28-kDa gene probe. St. Vincent, Wakulla, and Liberty isolates were similarly analyzed.

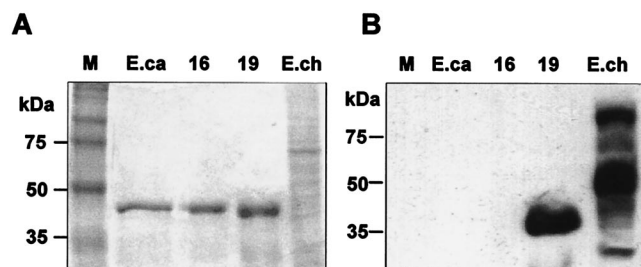


FIG. 8. Immunoblot analysis. Purified recombinant proteins for genes 16 and 19 of the Arkansas isolate, a recombinant antigen of a homologous gene from *E. canis* (ORF1) (24), and a whole-cell *E. chaffeensis* Arkansas isolate protein fraction were tested by Western blot analysis by using serum obtained from a B6 mouse immunized with live *E. chaffeensis* Arkansas isolate. (A) Coomassie blue stained gel; (B) corresponding immunoblot. Lanes: M, protein markers; E.ca, *E. canis* recombinant antigen; 16 and 19, the recombinant antigens made from *E. chaffeensis* isolate Arkansas 28-kDa OMP genes 16 and 19, respectively; E.ch, whole-cell antigen of *E. chaffeensis* isolate Arkansas.

ation in the paralogous genes was higher among each isolate compared to that found among orthologous genes from different isolates. These observations suggest that the paralogous genes in the genome may have generated by gene duplication, but since then they may have evolved independently from each other.

The loss or gain of two genes within the 28-kDa OMP locus is a novel observation that, to our knowledge, we report here for the first time. This finding supports the hypothesis that the locus undergoes major rearrangements, such as gene duplication or elimination, resulting in the loss or gain of a complete gene(s). The presence of gene 18 only in Group I isolates and gene 15a only in Group III isolates is evidence for this type of genetic change. Size variations in the 120-kDa and VLPT genes among different isolates also suggest that the insertion or deletion mutations of large pieces of DNA is a common occurrence in this bacterium. Variation in the number of repeats in these two loci may also have resulted from slippage mutations. The molecular basis and functional significance of the

observed genomic changes resulting from the gene deletion or insertions and variable number of repeats within protein coding sequences remains to be investigated.

Our RT-PCR analysis of total RNA isolated from cultured *E. chaffeensis* isolates revealed differences in the transcriptional activity of the locus. Expressed transcripts remained constant for two older isolates (Arkansas and St. Vincent) (4, 23) but are variable with the source of RNA for two relatively new isolates (Wakulla and Liberty) (31). To examine if mRNA expression coincides with protein synthesis, we also analyzed antigen expression by Western blot analysis. There was a perfect agreement between the transcribed and nontranscribed genes for the two genes tested in this study.

RT-PCR analysis of the paralogs of the 28-kDa genes of *E. chaffeensis* also were reported recently for the Arkansas isolate (15, 34). A locus highly homologous to the *E. chaffeensis* 28-kDa OMP locus from *E. canis* also was examined for gene expression by RT-PCR (19). According to Long et al. (15), RT-PCR analysis of RNA obtained from in vitro-cultured *E. chaffeensis* suggests the expression of 16 of the 22 genes. Ohashi et al. (19) reported the expression of all 22 genes from the *E. canis* locus. Our study on the transcriptional analysis of RNA spanning seven 3'-end clusters of OMP genes of in vitro-cultured *E. canis* (referred to as an α region [19]) revealed the expression of only three of seven genes (Ganta and Cheng, unpublished results). Only one OMP gene transcript of *E. canis* grown in a tick host is reported, and a higher level of expression of the same gene transcript is also reported for *E. canis* cultured at 25°C (32). Our present study analyzed RNA for the six tandemly arranged, paralogous genes (α region of the locus [19]) of a Group I isolate, Arkansas, Group II isolates St. Vincent and Wakulla, and a Group III isolate, Liberty. Our data were identical to those reported by Long et al. (15) for genes 15, 17, 18, and 19 of the Arkansas isolate. While we observed expression of gene 14 and no expression of gene 16, Long et al. (15) reported the opposite. Similarly, our unpublished RT-PCR data for *E. canis* differed from those reported by Ohashi et al. (19). Our data in the present study for Wakulla and Liberty isolates are variable in different preparations of

TABLE 3. Expression of the 28-kDa OMP analyzed by RT-PCR

Isolate analyzed (Group)	No. of times observed	Genes analyzed by RT-PCR	Genes positive for amplicons ^a
Arkansas (I)	3	14, 15, 16, 17, 18, 19	14, 15, 18, 19
St. Vincent (II)	3	14, 15, 16, 17, 19	14, 15, 17, 19
Wakulla (II)	1	14, 15, 16, 17, 19	15, 19
	1	14, 15, 16, 17, 19	15, 16, 17, 19
	1	14, 15, 16, 17, 19	15, 16, 17, 19
	3	14, 15, 16, 17, 19	14, 15, 16, 17, 19
Liberty (III)	1	14, 15a, 15, 16, 17, 19	15, 16, 19
	1	14, 15a, 15, 16, 17, 19	16, 17, 19
	1	14, 15a, 15, 16, 17, 19	14, 15a, 15, 16, 17
	2	14, 15a, 15, 16, 17, 19	14, 15a, 15, 16, 17, 19

^a If the amplified products were different for an isolate when analyzed by using three different batches of RNA made from cultures harvested on different culture days, the experiment was repeated two more times.

RNA. Despite the presence of 22 genes in the locus, and independent of the variations in the identified expressed genes reported by different laboratories, it is noteworthy that multiple genes are transcriptionally active. It is not clear why the expressed transcripts vary. There may be multiple reasons for the observed differences in the transcribed genes. Two possible reasons are (i) RNA isolated from cultured organisms may represent RNA derived from bacteria in a nonsynchronized state, i.e., obtained from a mix of different life cycle stages, and (ii) expression may be influenced by the culture conditions. The variability in gene expression at the transcript level raises questions about using RT-PCR data as a measure for examining and interpreting data for translation of the gene products.

Interestingly, despite differences in the RNA expression and the documentation of multiple transcripts, the detectable protein made from the 28-kDa OMP locus reported in the literature remained constant (15, 21). On the basis of the N-terminal amino acid sequence of expressed proteins from cultured *E. chaffeensis*, Ohashi et al. (21) and Long et al. (15) identified only one protein, i.e., the product of gene 19. Our present study also supports the expression of this protein. These observations suggest that while the data on the transcriptional activity are conflicting, there is consistency in the translated protein encoded from the 28-kDa OMP locus. More importantly, there is no evidence that supports the translation of more than one gene.

Long et al. (15) suggested that several genes may be transcribed but that not all genes are translated. An alternative explanation for the observed presence of multiple transcripts made from the 28-kDa locus is that they represent transcripts made for overlapping genes that may be present within the 28-kDa OMP gene locus. In overlapping genes, parts of the same DNA region can encode for more than one protein but can use different reading frames (27). Sharing the same genomic region to encode for two or more proteins is commonly reported for viruses, bacteria, and protozoans (1, 13, 29). To examine if overlapping genes may exist in *E. chaffeensis*, predicted amino acid sequences having open reading frames longer than 20 amino acids in the remaining two forward frames and the three reverse frames were identified for genes 14 to 19 of the Arkansas isolate 28-kDa locus and were subjected to BLAST homology search. Predicted amino acid sequences of several ORFs exhibited significant similarities (>50% identity) with known sequences. A few examples are listed in Fig. 9. These observations raise the possibility of the presence of overlapping genes in the 28-kDa OMP locus. The presence of transcripts as judged by positive RT-PCR products for several nontranslating 28-kDa genes, therefore, may represent transcripts derived from the possible overlapping genes located within the 28-kDa OMP locus.

If *E. chaffeensis* undergoes rapid genetic and antigenic changes to evade host responses during the course of infection in vertebrate hosts or in ticks, as was proposed earlier (24), differences in the genomic DNA and expressed antigens in all of the isolates established from naturally infected hosts would be expected. The data presented in this study for the 28-kDa locus for isolates established from a nonreservoir, accidental host (human) revealed the presence of deletion or insertion mutations spanning large segments of DNA. Sequence characterization of the 120-kDa OMP and VLPT genes also re-

Gene 16: forward open reading frame 2: homology: potato virus Y polyprotein (GenBank no.: X97895) identities = 50%, positives = 68%

```
LVVIS--TSVESICQALRILEFFLPRK--KEIQ--QLE
L+VIS TSV ++CQA+ IL+ RK KEI+ Q+E
LLVISMLTSVVAVCQAI-ILD---QRKYRKEIELMQIE
```

Gene 17: forward open reading frame 2: homology: BH4040 [*Bacillus halodurans*] (GenBank no.: AP001520); identities = 53%, positives = 56%

```
ICQVLRILA-CFLPKKKK-ILLHLCM
IC LA CF K+KK ILL CM
IC----LAFCFQAKRKAILL--CM
```

Gene 18: forward open reading frame 2: homology to NADH dehydrogenase [*Dinodon semicarinatus*] (GenBank no.: AB008539) identities = 50%, positives = 70%

```
FLMQYRTTMLVVVISISVGNMYQVF
FL+ TMLV+I+ + NMYQ+ F
FLI----TMLVITAN--NMYQLF
```

Gene 19: reverse open reading frame 3: homology to hypothetical protein T01G.6 [*Caenorhabditis elegans*] (GenBank no.: AF022978); identities = 42%, positives = 48%

```
EHN---IYVLH-SY-----NCYLDFLHQMFHK---ILQAL-----FLD
EHN I L SY NC LDF QMF+K IL+ + FLD
EHNSSKGTI--LDVSYLIVKAHNC-LDFPPQMFNKNLTIIEKMQCANEFLD
```

FIG. 9. GenBank search analysis of long ORFs other than 28-kDa OMP coding sequences. Predicted amino acid sequences of ORFs longer than 20 amino acids from forward frames 2 and 3 and the three reverse frames were subjected to a homology search with sequences in the databases. ORFs having significant sequence similarities (>50% identity) to sequence motifs of known sequences available in the GenBank database are identified. One example each from genes 16, 17, 18, and 19 are shown.

vealed size variations, resulting from the loss or gain of long, direct repeats in the coding sequences. There is no overlap between the genetic grouping based on the 28-kDa locus and the other loci. These nonoverlapping major genomic differences in different regions of the genome may have occurred to counter the host immune response in the reservoir host where *E. chaffeensis* infection can persist. Alternatively, heterogeneity observed in different genomic regions could result from independently evolved genes. These hypotheses can be tested by evaluating *E. chaffeensis* isolates obtained from a reservoir host.

At this time, it is unclear why *E. chaffeensis* has several tandemly arranged genes having only one expressed antigen. The significance of the variable number of repeats in 120-kDa and VLPT genes among isolates is also unclear. The presence of a multigene locus and the variable multirepeat sequences in *E. chaffeensis* are similar to some outer surface protein gene loci that have been shown to play important roles in immune evasion in the genomes of the pathogenic bacteria *B. burgdorferi*, *N. gonorrhoeae*, and *M. hominis* (12, 28, 36, 37). The functional significance of these gene loci in *E. chaffeensis*, therefore, requires the analysis of organisms recovered over time from a reservoir host that is infected by tick bite.

In conclusion, we identified three distinct genetic groups of *E. chaffeensis* as judged by the analysis of the 28-kDa OMP locus. Novel gene deletion or insertion mutations resulting in the loss or gain of genes were identified in this locus that allowed separation of isolates into three groups. The isolates within each group may represent separate *E. chaffeensis* strains. Isolates from each group were associated with severe or fatal disease, and so it appears that multiple pathogenic strains of this bacterium exist. There were also other genomic differences

that resulted from the loss or gain of long repeats within the coding sequences of 120-kDa and VLPT genes. These differences did not overlap with the genetic grouping of isolates established from the analysis of the 28-kDa OMP locus. The molecular heterogeneity in *E. chaffeensis* may have developed as a mechanism to evade the host immune response, or it could be the result of independently evolved genes. Genomic differences in the *E. chaffeensis* isolates may influence disease pathogenesis, a hypothesis that remains to be tested.

ACKNOWLEDGMENTS

This study was supported by the National Institutes of Health grants AI50785 and RR017686 and Kansas State University Agricultural Experimental Station Animal Health funds section 1433 grant 4-81321.

REFERENCES

- Behrens, M., J. Sheikh, and J. P. Nataro. 2002. Regulation of the overlapping pic/set locus in *Shigella flexneri* and enteroaggregative *Escherichia coli*. *Infect. Immun.* **70**:2915–2925.
- Breitschwerdt, E. B., B. C. Hegarty, and S. I. Hancock. 1998. Sequential evaluation of dogs naturally infected with *Ehrlichia canis*, *Ehrlichia chaffeensis*, *Ehrlichia equi*, *Ehrlichia ewingii*, or *Bartonella vinsonii*. *J. Clin. Microbiol.* **36**:2645–2651.
- Chen, S. M., V. L. Popov, H. M. Feng, and D. H. Walker. 1996. Analysis and ultrastructural localization of *Ehrlichia chaffeensis* proteins with monoclonal antibodies. *Am. J. Trop. Med. Hyg.* **54**:405–412.
- Dawson, J. E., B. E. Anderson, D. B. Fishbein, C. Y. Sanchez, C. Y. Goldsmith, K. H. Wilson, and C. W. Duntley. 1991. Isolation and characterization of an *Ehrlichia* sp. from a patient diagnosed with human ehrlichiosis. *J. Clin. Microbiol.* **29**:2741–2745.
- Dawson, J. E., K. L. Biggie, C. K. Warner, K. Cookson, S. Jenkins, J. F. Levine, and J. G. Olson. 1996. Polymerase chain reaction evidence of *Ehrlichia chaffeensis*, an etiologic agent of human ehrlichiosis, in dogs from southeast Virginia. *Am. J. Vet. Res.* **57**:1175–1179.
- Dawson, J. E., J. E. Childs, R. K. Biggie, C. Moore, D. Stallknecht, J. Shaddock, J. Bouseman, E. Hofmeister, and J. G. Olson. 1994. White-tailed deer as a potential reservoir of *Ehrlichia* spp. *J. Wildl. Dis.* **30**:162–168.
- Devereux, J. 1984. Genetics computer group sequence analysis software package, version 6.1. *Nucleic Acids Res.* **12**:387–395.
- Dugan, V. G., S. E. Little, D. E. Stallknecht, and A. D. Beall. 2000. Natural infection of domestic goats with *Ehrlichia chaffeensis*. *J. Clin. Microbiol.* **38**:448–449.
- Dumler, J. S., A. F. Barbet, C. P. Bekker, G. A. Dasch, G. H. Palmer, S. C. Ray, Y. Rikihisa, and F. R. Rurangirwa. 2001. Reorganization of genera in the families Rickettsiaceae and Anaplasmataceae in the order Rickettsiales: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila*. *Int. J. Syst. Evol. Microbiol.* **51**:2145–2165.
- Fishbein, D., L. Sawyer, C. Holland, E. Hayes, W. Okoroanyanwu, B. Williams, R. Sikes, M. Ristic, and J. McDade. 1987. Unexplained febrile illnesses after exposure to ticks: infection with an *Ehrlichia*? *JAMA* **257**:3100–3104.
- Ganta, R. R., M. J. Wilkerson, C. Cheng, A. M. Rokey, and S. K. Chapes. 2002. Persistent *Ehrlichia chaffeensis* infection occurs in the absence of functional major histocompatibility complex class II genes. *Infect. Immun.* **70**:380–388.
- Haas, R., and T. F. Meyer. 1986. The repertoire of silent pilus genes in *Neisseria gonorrhoeae*: evidence for gene conversion. *Cell* **44**:107–115.
- Iwabe, N., and T. Miyata. 2001. Overlapping genes in parasitic protist *Giardia lamblia*. *Gene* **280**:163–167.
- Kordick, S. K., E. B. Breitschwerdt, B. C. Hegarty, K. L. Southwick, C. M. Colitz, S. I. Hancock, J. M. Bradley, R. Rumbough, J. T. McPherson, and J. N. MacCormack. 1999. Coinfection with multiple tick-borne pathogens in a Walker Hound kennel in North Carolina. *J. Clin. Microbiol.* **37**:2631–2638.
- Long, S. W., X. F. Zhang, H. Qi, S. Standaert, D. H. Walker, and X. J. Yu. 2002. Antigenic variation of *Ehrlichia chaffeensis* resulting from differential expression of the 28-kilodalton protein gene family. *Infect. Immun.* **70**:1824–1831.
- 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.
- Maniatis, T., E. F. Fritsch, and J. Sambrook. 1982. Molecular cloning: a laboratory manual. Cold Spring Harbor Laboratory, Cold Spring Harbor, N.Y.
- McBride, J. W., X. Yu, and D. H. Walker. 1999. Molecular cloning of the gene for a conserved major immunoreactive 28-kilodalton protein of *Ehrlichia canis*: a potential serodiagnostic antigen. *Clin. Diagn. Lab. Immunol.* **6**:392–399.
- Ohashi, N., Y. Rikihisa, and A. Unver. 2001. Analysis of transcriptionally active gene clusters of major outer membrane protein multigene family in *Ehrlichia canis* and *E. chaffeensis*. *Infect. Immun.* **69**:2083–2091.
- Ohashi, N., A. Unver, N. Zhi, and Y. Rikihisa. 1998. Cloning and characterization of multigenes encoding the immunodominant 30-kilodalton major outer membrane proteins of *Ehrlichia canis* and application of the recombinant protein for serodiagnosis. *J. Clin. Microbiol.* **36**:2671–2680.
- Ohasi, N., N. Zhi, Y. Zhang, and Y. Rikihisa. 1998. Immunodominant major outer membrane proteins of *Ehrlichia chaffeensis* are encoded by a polymorphic multigene family. *Infect. Immun.* **66**:132–139.
- Paddock, C. D., D. P. Suchard, K. L. Grumbach, W. K. Hadley, R. L. Kerschmann, N. W. Abbey, J. E. Dawson, B. E. Anderson, K. G. Sims, J. S. Dumler, and B. G. Herndier. 1993. Brief report: fatal seronegative ehrlichiosis in a patient with HIV infection. *N. Engl. J. Med.* **329**:1164–1167.
- 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.
- Reddy, G. R., and C. P. Streck. 1999. Variability in the 28-kDa surface antigen protein multigene locus of isolates of the emerging disease agent *Ehrlichia chaffeensis* suggests that it plays a role in immune evasion. *Mol. Cell Biol. Res. Commun.* **1**:167–175.
- Reddy, G. R., C. R. Sulsona, A. F. Barbet, S. M. Mahan, M. J. Burrige, and A. R. Alleman. 1998. Molecular characterization of a 28 kDa surface antigen gene family of the tribe Ehrlichiae. *Biochem. Biophys. Res. Commun.* **247**:636–643.
- Reddy, G. R., C. R. Sulsona, R. H. Harrison, S. M. Mahan, M. J. Burrige, 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.
- Rogozin, I. B., A. N. Spiridonov, A. V. Sorokin, Y. I. Wolf, I. K. Jordan, R. L. Tatusov, and E. V. Koonin. 2002. Purifying and directional selection in overlapping prokaryotic genes. *Trends Genet.* **18**:228–232.
- Segal, E., P. Hagblom, H. S. Seifert, and M. So. 1986. Antigenic variation of gonococcal pilus involves assembly of separated silent gene segments. *Proc. Natl. Acad. Sci. USA* **83**:2177–2181.
- Shmulevitz, M., Z. Yameen, S. Dawe, J. Shou, D. O'Hara, I. Holmes, and R. Duncan. 2002. Sequential partially overlapping gene arrangement in the tricistronic S1 genome segments of avian reovirus and Nelson Bay reovirus: implications for translation initiation. *J. Virol.* **76**:609–618.
- Standaert, S. M., T. Yu, M. A. Scott, J. E. Childs, C. D. Paddock, W. L. Nicholson, J. J. Singleton, and M. J. Blaser. 2000. Primary isolation of *Ehrlichia chaffeensis* from patients with febrile illnesses: clinical and molecular characteristics. *J. Infect. Dis.* **181**:1082–1088.
- Sumner, J. W., J. E. Childs, and C. D. Paddock. 1999. Molecular cloning and characterization of the *Ehrlichia chaffeensis* variable-length PCR target: an antigen-expressing gene that exhibits interstrain variation. *J. Clin. Microbiol.* **37**:1447–1453.
- Unver, A., N. Ohashi, T. Tajima, R. W. Stich, D. Grover, and Y. Rikihisa. 2001. Transcriptional analysis of p30 major outer membrane multigene family of *Ehrlichia canis* in dogs, ticks, and cell culture at different temperatures. *Infect. Immun.* **69**:6172–6178.
- Walker, D. H., and J. S. Dumler. 1997. Human monocytic and granulocytic ehrlichioses. Discovery and diagnosis of emerging tick-borne infections and the critical role of the pathologist. *Arch. Pathol. Lab. Med.* **121**:785–791.
- Yu, X., J. W. McBride, X. Zhang, and D. H. Walker. 2000. Characterization of the complete transcriptionally active *Ehrlichia chaffeensis* 28 kDa outer membrane protein multigene family. *Gene* **248**:59–68.
- Yu, X. J., P. Crocquet-Valdes, and D. H. Walker. 1997. Cloning and sequencing of the gene for a 120-kDa immunodominant protein of *Ehrlichia chaffeensis*. *Gene* **184**:149–154.
- Zhang, J., J. M. Hardham, A. G. Barbour, and S. J. Norris. 1997. Antigenic variation in Lyme disease *Borreliae* by promiscuous recombination of VMP-like sequence cassettes. *Cell* **89**:275–285.
- Zhang, Q., and K. S. Wise. 1996. Molecular basis of size and antigenic variation of a *Mycoplasma hominis* adhesin encoded by divergent vaa genes. *Infect. Immun.* **64**:2737–2744.