Structure of the Expression Site Reveals Global Diversity in MSP2 (P44) Variants in *Anaplasma phagocytophilum*^{▽†}

Anthony F. Barbet,^{1*} Anna M. Lundgren,¹ A. Rick Alleman,² Snorre Stuenkel,³ Anneli Bjöersdorff,⁴ Richard N. Brown,⁵ Niki L. Drazenovich,⁶ and Janet E. Foley⁶

Department of Infectious Diseases and Pathology¹ and Department of Physiological Sciences,² University of Florida, Gainesville, Florida 32611; Norwegian School of Veterinary Science, Sandnes N-4325, Norway³; Department of Clinical Microbiology, Kalmar County Hospital, Kalmar, Sweden⁴; Department of Wildlife, Humboldt State University, Arcata, California 95521⁵; and Department of Medicine and Epidemiology, School of Veterinary Medicine, University of California, Davis, California 95616⁶

Received 19 May 2006/Returned for modification 26 June 2006/Accepted 30 August 2006

Anaplasma phagocytophilum, a recently reclassified bacteria in the order *Rickettsiales*, infects many different animal species and causes an emerging tick-borne disease of humans. The genome contains a large number of related genes and gene fragments encoding partial or apparently full-length outer membrane protein MSP2 (P44). Previous data using strains isolated from humans in the United States suggest that antigenic diversity results from RecF-mediated conversion of a single MSP2 (P44) expression site by partially homologous donor sequences. However, whether similar mechanisms operate in naturally infected animal species and the extent of global diversity in MSP2 (P44) are unknown. We analyzed the structure and diversity of the MSP2 (P44) expression site in strains derived from the United States and Europe and from infections of different animal species, including wildlife reservoirs. The results show that a syntenic expression site is present in all strains of *A. phagocytophilum* investigated. This genomic locus contained diverse MSP2 (P44) variants in all infected animals sampled, and variants also differed at different time points during infection. Although similar variants were found among different populations of U.S. origin, there was little sequence identity between U.S. strain variants (including genomic copies from a completely sequenced U.S. strain) and expression site variants infecting sheep and dogs in Norway and Sweden. Finally, the possibility that combinatorial mechanisms can generate additional diversity beyond the basic donor sequence repertoire is supported by the observation of shared sequence blocks throughout the MSP2 (P44) hypervariable region in reservoir hosts. These data suggest similar genetic mechanisms for *A. phagocytophilum* variation in all hosts but worldwide diversity of the MSP2 (P44) outer membrane protein.

Anaplasma phagocytophilum is the causative agent of an emerging tick-borne zoonosis, human granulocytic anaplasmosis. The organism infects many species worldwide, including those of rodents, deer, bears, horses, cattle, dogs, and sheep (12). It has been known to cause persistent disease in domestic ruminants (e.g., tick-borne fever in sheep and pasture fever in cattle) for many years (34). In humans, infections are thought to be normally acute, although long-term adverse health outcomes have been recognized even after antibiotic treatment (36). The closely related *Anaplasma marginale* also causes persistent infections that can last for the lifetime of the host, and the mechanisms of persistence have been well studied recently. Despite a small genome of only 1.2 Mb, *A. marginale* can express multiple variants of two surface proteins, MSP2 and MSP3, during infection (8, 16, 19, 20). The mechanism appears similar for both MSP2 and MSP3, in that each protein has a unique genomic expression site (3, 31). The *msh2* and *msh3* expression sites both encode polycistronic mRNA species and these expression sites frequently undergo segmental gene con-

version by sequences, termed functional pseudogenes, that are duplicated from *msh2* or *msh3* copies elsewhere in the genome (7, 9, 31). This leads to an extensive assortment of different surface proteins. The majority of changes produced by gene conversion are in a central hypervariable region of MSP2 or in a longer MSP3 hypervariable region. The hypervariable region of MSP2 contains exposed surface epitopes that are recognized by antibody subsequent but not prior to the appearance of the respective variants in the bloodstream (19). Important recent observations are, first, that the mosaic complexity of the *msh2* and *msh3* expression sites increases during infection, i.e., fewer segmental changes are required to explain expression site structure early in infection and more are required as infection progresses (21, 31) and, second, when superinfection with a different strain occurs naturally, the strains that superinfect possess *msh2* pseudogene repertoires different from those of strains that do not (37). This suggests that the ability of *A. marginale* to persist in individual infections and to cause disease in cattle herds may be related to the extent of combinatorial diversity that can be created in these surface proteins.

Although *A. phagocytophilum* has been less well studied, there are some similarities between the two organisms. *A. phagocytophilum* also expresses a major surface protein that is an ortholog of MSP2, which we will term here MSP2 (P44) (25, 33, 40). Like MSP2 of *A. marginale*, MSP2 (P44) contains a central hypervariable region and is expressed from a genomic expression site subject to frequent gene conversion by sequence duplicated from

* Corresponding author. Mailing address: Department of Infectious Diseases and Pathology, College of Veterinary Medicine, Box 110880, Gainesville, FL 32611-0880. Phone: (352) 392-4700, ext. 5819. Fax: (352) 392-9704. E-mail: barbeta@mail.vetmed.ufl.edu.

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

[▽] Published ahead of print on 11 September 2006.

msp2 (*p44*) copies elsewhere in the genome (4, 27). The promoter regions in the two expression sites have nearly identical sequences (2). There are also differences between the two systems. In the *A. phagocytophilum* genome, there are many more *msp2* (*p44*) pseudogene copies, about 100 compared to about 7 in *A. marginale* (6, 14). Also, in the variants analyzed to date, there appears to be less use of small donor gene segments to create mosaic complexity in *A. phagocytophilum* than that in *A. marginale*. In fact, the expression site hypervariable regions have closely resembled those in the donor copies, except at the borders of the converted segments (4, 26, 27). However, as yet, relatively few expression site variants of *A. phagocytophilum* have been analyzed, and these have all been derived from human strains. The purpose of this study was to investigate the structure and diversity of the *msp2* (*p44*) expression site in a greater range of animal species in order to ascertain whether similar persistence mechanisms could operate in the full range of infections caused by *A. phagocytophilum*.

MATERIALS AND METHODS

Origin of strains of *A. phagocytophilum*. DNA was extracted from six human blood samples that were submitted to the New York State Department of Health and that were PCR positive for *A. phagocytophilum*, as has been described previously (4). In vitro culture and growth of strains of *A. phagocytophilum* originally isolated from infected human patients from New York state (NY-18 and NY-37) and Minnesota (HGE2) have also been reported previously (1, 22, 24, 29). *A. phagocytophilum* infections have been identified by PCR of DNA extracted from the blood of dusky-footed wood rats, which are suspected reservoir hosts of the disease, in California and from Californian black bears (17). For this study, we analyzed the populations of organisms present in the blood of six wood rats trapped at the same field site in the Hoopa Valley, Humboldt County, California, in January (wood rats 1, 2, and 6) or March (wood rats 3, 4, and 5) 2003 and from one black bear that was also from the Hoopa Valley site. DNA was isolated from the blood of three dogs that were clinically ill with granulocytic anaplasmosis in Sweden and from two sheep experimentally infected with a cryopreserved isolate of *A. phagocytophilum* from the blood of a Norwegian flock with tick-borne fever. DNA was analyzed from the experimentally infected sheep at 5 days postinfection and also for sheep 1 at 17 days postinfection.

Determination of the structure of the *msp2* (*p44*) expression site. The sequence of the *msp2* (*p44*) expression site in U.S. strains from humans has been defined previously (4). Initial attempts to amplify the *msp2* (*p44*) gene in the expression site using the oligonucleotide primers AB 1000 (CCGCTGAAGT GAGGAGACGA) and AB 1001 (AAGTACCGCAGGAAGTAGAAT), defined previously, were successful for all U.S. strains but unsuccessful with the European samples. We then used different combinations of primers and also a primer-walking strategy (DNA walking SpeedUp kit; Seegene USA, Rockville, MD) to obtain provisional sequences for the expression sites from the European strain. These sequences differed from the U.S. strain sequence, particularly in regions immediately flanking the *msp2* (*p44*) gene in the expression site, but did contain these conserved features: an upstream gene similar to *p44ESup1*, a nearly identical 5' promoter, and a 3'-truncated *recA* gene (Fig. 1). To confirm and correct this provisional sequence, PCR amplification with oligonucleotide primers AB 1207 (GGGAGTGCTCTGGTTAGATTAGG) and AB 1198 (GC GAGGAAGCAATGAGAATAG) was used to amplify a 3.1-kb fragment containing *p44ESup1*, *msp2* (*p44*), and truncated *recA*. This fragment was gel isolated, cloned in the plasmid vector pCR4-TOPO (Invitrogen, Carlsbad, CA), and sequenced on both strands. The promoter region, which has proven unstable and difficult to clone in previous studies (2), was PCR amplified separately with primers AB 1041 (ATGTCAGTACCGGCATATCTTGAAATC) and AB 1200 (GCATAGAACCCATCGGCTTCAC) to yield an overlapping 331-bp fragment that was sequenced directly without cloning.

Sequencing of *msp2* (*p44*) expression site variants. DNA from all samples of U.S. origin was amplified by PCR using oligonucleotide primers AB 1000 and AB 1001 to amplify a fragment of 1.4 kb containing the *msp2* (*p44*) gene expression site, as described previously (4). Due to the divergence of expression site sequence in European strains, it was necessary to design alternative primers based on the sequencing data obtained for these strains. DNA samples from dogs infected with *A. phagocytophilum* of Swedish origin were PCR amplified with oligonucleotide primers AB 1221 (ATAGAACAAGAGCAGGGAGAAGAC),

which anneals just 3' to the truncated *recA* gene, and AB 1222 (GTGTTGAAT CCGGACATATTGAAG), which anneals in the 5' region flanking the *msp2* (*p44*) gene in the expression site. These primers generated a fragment of approximately 1.9 kb. DNA from *A. phagocytophilum* infecting Norwegian sheep was amplified with oligonucleotide primers AB 1221 and AB 1227 (TCTGTCT TGGAGAGTATTGAGTC), which anneals in the 5' region flanking the *msp2* (*p44*) gene expression site in these strains, to generate a fragment of ~2 kb (Fig. 1). In all cases, PCR-amplified fragments were cloned into the plasmid vector pCR4-TOPO and individual colonies were grown overnight in 96-well deep blocks containing 1.5 ml LB medium and kanamycin (50 µg/ml). Plasmid DNA was prepared from cultures by centrifuging blocks at $1,100 \times g$ for 15 min, resuspending cells in 400 µl of 50 mM Tris-HCl, pH 8.0, 10 mM EDTA, 100 µg/ml RNase A, and then adding 400 µl of 200 mM NaOH, 1% sodium dodecyl sulfate and inverting five times to lyse cells. Four hundred microliters of 3 M sodium acetate, pH 4.5, was added, and the plates were covered with sealing tape and inverted five times. The blocks were placed at -80°C for 1 to 2 h and then thawed and centrifuged for 30 min at $2,830 \times g$ at 4°C . Nine hundred microliters of cleared supernatant was transferred to a 96-well filter plate (Unifilter; Whatman, Clifton, NJ) on a new 96-well block. An equal volume of isopropanol was added to each well, and the block was covered with sealing tape and inverted 1 to 2 times before being centrifuged for 45 to 60 min at $2,830 \times g$. The supernatant was carefully decanted, and pellets of DNA were washed with 1 ml 70% ethanol and air dried before resuspension in TE buffer (10 mM Tris-HCl, 1 mM EDTA, pH 8.0). All samples were digested with EcoRI to release insert DNA and with EcoRI and RsaI to analyze restriction fragment length polymorphism (RFLP) patterns. Digested DNA was resolved by electrophoresis on 1.5% agarose gels and visualized by staining with GelStar nucleic acid stain (Cambrex, Rockland, Maine). Clones were selected for DNA sequencing based on the different RFLP patterns obtained. Representative clones were selected for sequencing from each of the major RFLP patterns, excluding singlet patterns. The analyses described here included RFLP analysis from expression site clones (469 infected human, 342 HGE2 culture, 311 NY-18 culture, 36 Webster culture, 218 infected wood rat, 20 infected bear, 229 infected Swedish dog, and 67 infected Norwegian sheep) of *A. phagocytophilum*. Sequencing was performed at the University of Florida DNA Sequencing Core Laboratory (Gainesville, FL) by using ABI Prism dye terminator cycle sequencing protocols developed by Applied Biosystems (Foster City, California). The fluorescently labeled extension products were sequenced on both strands by using PerkinElmer/Applied Biosystems automated DNA sequencers. A minority of the sequences obtained contained small changes in amino acid sequence, such as a single amino acid substitution, or an apparent deletion that resulted in no open reading frame through the *msp2* (*p44*) hypervariable region. For the purpose of the analyses described here, we did not attempt to determine whether these changes resulted from differences in the original template DNA or were introduced as a result of PCR amplification, clone instability, or sequencing errors that were reproduced on both strands. However, an arbitrary decision was made to include in the alignments all variants with minor amino acid substitutions, so that identities with genome copies would not be missed. Likewise, we excluded from the analysis any expression site sequence without a complete open reading frame through the hypervariable region. Such sequences have been observed previously in the syntenic *A. marginale* expression site (5), however, and could be a natural result of frequent pseudogene recombination into the *msp2* (*p44*) expression site.

Sequence analyses and alignments. Nucleotide sequences were analyzed using the Wisconsin Package, version 10.3-Unix, (Accelrys, Inc., San Diego, CA), running under SunOS 5.8, available through the Biological Computing Core facilities of the Interdisciplinary Center for Biotechnology Research at the University of Florida. Sequence alignments were made initially using PILEUP, and similarities were displayed using PLOTSIMILARITY and PRETTYBOX. $K_a:K_s$ ratios (nonsynonymous substitutions per nonsynonymous site to synonymous substitutions per synonymous site) were determined with DIVERGE. To align multiple sequences from the sequenced HZ strain genome with expression site sequences (see Fig. S1 in the supplemental material), the annotated *msp2* (*p44*) genomic loci in GenBank CP000235 were first extracted as a multiple Fasta file using ARTEMIS (38) (<http://www.sanger.ac.uk/Software/Artemis/>) and aligned with DIALIGN (32), running on a Relion 130 linux server (Penguin Computing, San Francisco, CA). This effectively identified the presence of the conserved regions flanking the hypervariable region in genomic open reading frames of very different lengths. The genome and expression site sequences were then trimmed to the LAKT residues present on both sides of the hypervariable regions and aligned with PILEUP. The aligned multiple sequence file was visualized with CHROMA (23) (<http://www.lleu.org.uk/chroma>). To determine the relationships between the aligned sequences in terms of percent amino acid sequence identities, a Fasta file containing all trimmed genome and expression site se-

**NY-18 vs. Swdog
vs. Norsheep**

**NY-18 vs. NY-37
vs. HZ**

FIG. 1. *Msp2* expression site variability in three U.S. and two European strains of *A. phagocytophilum*. PLOTSIMILARITY comparison of U.S., Norwegian, and Swedish strains (top panel) or three U.S. strains (bottom panel). The positions of coding sequences for truncated RecA, MSP2 (P44), P44ESup1, and the promoter (P) are shown above the PLOTSIMILARITY profiles. A similarity score of 1.0 indicates identical sequence in a sliding window of 10 nucleotides, and a decreasing score from 1.0 to 0 indicates increasing variation. Sequences of U.S. strains NY-18, NY-37, and HZ are from GenBank AY164490, AY137510, and CP000235, respectively.

quences was analyzed with MatGAT (Matrix global alignment tool) (10). The percentage identities determined by MatGAT for an all-against-all comparison were exported to an Excel spreadsheet for final determination of those alignments meeting or exceeding cutoff values of 70 or 90% identity. The final alignments analyzed (see Fig. S1 in the supplemental material) contained 94 *A. phagocytophilum* genomic loci of *msp2 (p44)*, 74 expression site variants of U.S. origin (4 bear, 19 wood rat, 25 cultured human origin, and 26 human), 19 Swedish dog variants, and 23 Norwegian sheep variants.

Nucleotide sequence accession numbers. The nucleotide sequences reported here have GenBank accession numbers DQ519449 through DQ519570. The *msp2 (p44)* variable region sequences present in the expression site (see Fig. S1

in the supplemental material) are DQ519449 through DQ519564, sequences of the *msp2 (p44)* expression site in the Norwegian sheep and Swedish dog strains are DQ519565 and DQ519566, and those of complete *msp2 (p44)* genes in the expression site (variants B1v1, D2v1, Sh1t1v4, and WR1v2) are DQ519567 through DQ519570.

RESULTS

Structure of an *msp2 (p44)* expression site in European strains of *A. phagocytophilum*. We and others (4, 27) have

described the structure of an expression site for the major RNA species encoding MSP2 (P44) in strains of *A. phagocytophilum* derived from human infections in the United States. This expression site encodes a polycistronic RNA and is polymorphic because of frequent gene conversion by functional pseudogenes duplicated from elsewhere in the genome. Consequently, the RNA found at a single time point during a single human infection is also polymorphic and encodes diverse MSP2 (P44) species having different hypervariable regions but similar N and C termini. This expression site is flanked upstream by a promoter similar to that of the syntenic *msp2* expression site in *A. marginale* and downstream by a truncated *recA* gene. Initial experiments investigated the structure of this expression site in European strains of *A. phagocytophilum* derived from strains infecting sheep and dogs in Norway and Sweden, respectively. Oligonucleotide primers designed to flank *msp2* (*p44*) in the expression sites of U.S. strains (4) failed to amplify a product from the European strains, suggesting differences in the sequences of these flanking regions. Subsequent amplifications using different sets of flanking primers did amplify products of the expected size and these products were sequenced and extended in both 5' and 3' directions. Features of the final sequences that were obtained from the European strains suggested that they contained the expression site syntenic to that found previously in U.S. strains of *A. phagocytophilum*. The DNA sequences were similar to those of U.S. strains in that the MSP2 (P44) coding region was flanked upstream by a sequence nearly identical to the promoter region in U.S. strains and downstream by a gene encoding a truncated RecA (Fig. 1). Also, this site encoded a polymorphic MSP2 (P44) with different central hypervariable regions present in organisms sampled at a single time point in a single infection (see Fig. S1 in the supplemental material). Upstream of the gene encoding MSP2 (P44), but downstream of the promoter in the expression site, was a second coding region with best Blastp match (2e-131) to the equivalent gene, *p44ESup1* (for details, see reference 4), also known as *omp-1n* (for details, see reference 27) in U.S. strains. A comparison of the sequence of this presumptive expression site from European strains with the available sequence from U.S. strains revealed considerably greater diversity than that seen among U.S. strains alone (Fig. 1). The sequence differences were particularly noteworthy in the intergenic regions flanking the *msp2* (*p44*) gene, hence explaining the previous failure to PCR amplify this genomic expression site using oligonucleotide primers derived from the sequence of U.S. strains.

Figure 2 compares the amino acid sequences of P44ESup1 and MSP2 (P44) from U.S. and European strains of *A. phagocytophilum* and from strains infecting different animal hosts. Overall, the structures of P44ESup1 and MSP2 (P44) were well conserved. The sequence of P44ESup1 derived from a strain infecting Norwegian sheep differed by 53 amino acid substitutions from a strain infecting humans in the United States. These changes were distributed throughout the molecule. The strain of *A. phagocytophilum* from infected dogs in Sweden differed by 16 amino acid substitutions from the U.S. sequence. Despite these amino acid sequence changes, the K_a/K_s ratios for the P44ESup1 nucleotide sequences were 0.179 to 0.272, which may suggest evolutionary selection pressure for conservation of this gene.

A comparison of the MSP2 (P44) sequences revealed that, despite the differences in flanking regions (Fig. 1), the N and C termini of the coding region itself were very similar in U.S. and European strains (Fig. 2B). To extend this comparison further, the MSP2 (P44) expression sites from *A. phagocytophilum* infecting diverse species (humans, sheep, dogs, wood rats, and bears) were also compared. With the exception of the central hypervariable region, all MSP2 (P44) sequences were similar and few changes were observed that were unique to the infections of a particular species. One possible exception was the N-terminal sequence MKRGRLL, which was found in MSP2 (P44) from both Swedish dog and Norwegian sheep strains but was MRKGKII in the human, bear, and wood rat strains from the United States.

Diversity in the central hypervariable region of MSP2 (P44) among strains. The previous data demonstrated that, like U.S. strains of *A. phagocytophilum* derived from human infections, *A. phagocytophilum* strains from Europe and strains infecting diverse animal species differed primarily in a region of the *msp2* (*p44*) expression locus encoding about 120 amino acids in the middle of the MSP2 (P44) protein (Fig. 2B). To examine the diversity of this region more closely, independent clones of the *msp2* (*p44*) expression site were prepared and sequenced from all *A. phagocytophilum* strains. The central hypervariable regions encoding MSP2 (P44) and flanking conserved sequences from all clones were aligned (see Fig. S1 in the supplemental material). We also included hypervariable region sequences from our previous studies of this expression site, allowing the alignment of 116 total sequences derived from U.S. human infections and human strains grown in culture, as well as from the *A. phagocytophilum* strains infecting European dogs and sheep and U.S. bears and wood rats. The alignment showed the same flanking conserved regions to be present in most variants and that many different variants were present in each population analyzed. These variants were generally different between individual animals and even in the same animal when samples were available at multiple time points of a single infection. These data suggest (i) that similar antigenic variation mechanisms through expression site recombination are available to all strains of *A. phagocytophilum* and (ii) that there is great diversity in the potential repertoire of expressed sequences. The alignment also reveals some of the structural constraints on this variation, which include amino acids with conserved general structural properties, such as polarity and size, surrounding the highly conserved cysteine, phenylalanine, and tryptophan residues (see consensus sequence line in Fig. S1 in the supplemental material) recognized previously (29).

We employed MatGAT (10) to analyze the extent of expression site diversity. MatGAT generates pairwise identity matrices from an all-against-all comparison of the 116 aligned sequences. Excluding self-against-self comparisons, this MatGAT analysis included a comparison of 6,670 individual pairwise alignments having an average of 50.5% identity over the region analyzed (LAKT to LAKT). Of these 6,670 alignments, 68 (1%) had >90% identity and 106 (1.6%) had >70% identity. The small number of similar alignments suggests enormous potential global diversity in the hypervariable region sequences of expressed MSP2 (P44). Of the 68 alignments with >90% identity, 51 were among variants from different U.S. origin populations, 14 were among variants from

A

swdog	MVFRGLRCLF	FTCCLLVSFC	LVERGEADGF	YASLGWGESV	ADFGVEHLKV	KGITGSVMGP	60
ushuman	MVFGSLRCLF	FTYCLLVSPC	LVERGEAGGF	YASLGWGESV	VDFGVEHLKV	RGITGSVMGP	60
norsheep	MVFRGLRCLF	FTCCLLVPFS	LVERGEADGF	YASLGWGESV	ADFGVEHLKV	KGINGGVMGP	60
swdog	RKRRELDVAVKY	GHDTFSGKIG	GVGSGEFGDK	YRPVYTEDFG	LSGSFGYRFG	NLGIEIEGSE	120
ushuman	RKSELDAVKY	GHDTFSGKIG	GVGSGEFGDK	YRPVYTEDFG	LSGSFGYRFG	NLGIEIEGSE	120
norsheep	RKRELDVAVRY	RHDTFARGLLG	GVGSGEFGDK	YRPVYTEDFG	LSGSFGYRFG	NLGVEIEEGGE	120
swdog	QEFHPSAAGY	KIEGNAVHFA	FTTGKRGAKG	EVSGFVGENP	GMHLGLVLLN	GCYSGRLRLDE	180
ushuman	QEFHPSAAGY	KIEGNAVHFA	FTTGKSGAKG	EVSGFVGENP	GMHLGLVLLN	GCYSGRLRLDE	180
norsheep	REFHPSAAGY	KVEGNAMHFA	FTTGKRGAGE	KVSGLVGENS	GVHLGLVMLN	TCYSGRLRLDE	180
swdog	KNGIYTCLGL	GLAATDYLGG	LGRLRALWNA	KIGVELQFNK	NLAMFGEVYY	RGFGVLPQKV	240
ushuman	KNGIYSCCLGL	GLAATDYLGG	LGRLRALWNA	KIGVELQFTK	NLTMFGEVYY	RGFGVLRQKV	240
norsheep	RNGIYTCFGL	GLAAADYPGG	LGKRLRALWGA	KIGVELQFAK	NLAMFGEVYY	HGFQMSPGKM	240
swdog	RVFSVAAVDG	VDTLSSSGID	VSYVGNACGV	RYLF*274			
ushuman	RVFSVAAVEG	VDTLSSSGMN	VSYVGNACGV	RYLF*274			
norsheep	RVFSVAAATEG	VDTLSTGTID	ISYLGNGCGI	RYLF*274			

B

uswrat	MRKGGKIILGS	VMMSSMAIVMA	GNDVRAHDDV	SALDTGGAGY	FYVGLDYSPA	FSKIRDFSIR	60
ushuman	MRKGGKIILGS	VMMSSMAIVMA	GNDVRAHDDV	SALDTGGAGY	FYVGLDYSPA	FSKIRDFSIR	60
usbear	MRKGGKIILGS	VMMSSMAIVMA	GNDVRAHDDV	SALDTGGAGY	FYVGLDYSPA	FSKIRDFSIR	60
swdog	MKRGRLLLLGS	VMMSSMAIVMA	GNDVRAHDDV	SALDTGGAGY	FYVGLDYSPA	FSKIRDFSIR	60
norsheep	MKRGRLLLLGS	VMMSSMAIVMA	GNDVRAHDDV	SALDTGGAGY	FYVGLDYSPA	FSKIRDFSIR	60
uswrat	ESNGETKAVY	PYLKDGKSVK	LESHKFDWNT	PDPRIGFKDN	MLVAMEGSVG	YGIGGARVEL	120
ushuman	ESNGETKAVY	PYLKDGKSVK	LESHKFDWNT	PDPRIGFKDN	MLVAMEGSVG	YGIGGARVEL	120
usbear	ESNGETKAVY	PYLKDGKSVK	LESHKFDWNT	PDPRIGFKDN	MLVAMEGSVG	YGIGGARVEL	120
swdog	ESNGETKAVY	PYLKDGKSVK	LESHKFDWNT	PDPRIGFKDN	MLVAMEGSVG	YGIGGARVEL	120
norsheep	ESNGETKAVY	PYLKDGKSVK	LESHKFDWNT	PDPRIGFKDN	MLVAMEGSVG	YGIGGARVEL	120
uswrat	EIGYERFKTK	GIRDSGSKED	EADTVYLLAK	ELAYDVVTGQ	TDKLTAAALAK	TSGKDIVQFA	180
ushuman	EIGYERFKTK	GIRDSGSKED	EADTVYLLAK	ELAYDVVTGQ	TDNLAALAK	TSGKDIVQFA	180
usbear	EIGYERFKTK	GIRDSGSKED	EADTVYLLAK	ELAYDVVTGQ	TDNLAALAK	TSGKDIVQFA	180
swdog	EIGYERFKTK	GIRDSGSKED	EADTVYLLAK	ELAYDVVTGQ	TDNLAALAK	TSGKDIVQFA	180
norsheep	EIGYERFKTK	GIRDSGSKED	EADTVYLLAK	ELAYDVVTGQ	TDNLAALAK	TSGKDFVQFA	180
uswrat	KAVEISSPDI	DGKVCNNTHG	DRDKSHOAGT	A.TFGDKAQA	TV...AAQCS	GFATKKEGLT	236
ushuman	KTLNISHSNI	DGKVCCTGSHA	DLAGKTNAGK	KFVVNPPEASG	STDGDTSSQCS	GLG.HSSGVAT	239
usbear	NAVKISSPAI	DGKVCCKGSA	AV...GTAQK	EYA.RVPDDSD	SSAGKTAQCS	GL...NKSDAK	235
swdog	KAVGVSHPN	DKKVCNNGSHA	ALSESASAAA	DYA.SDFSSG	STTAKTAQCS	GF...KNSGHT	237
norsheep	KTVAIISHPN	NEKVCVTKDG	SSSKNKKTGYG	QYALS...TNN	SADHTVALCG	GAGHASNGSS	238
uswrat	...FSKFA	TEVKIIGE.KN	WPTGR...IYD	SSSDKEASON	SNAKAVATDL	TK.LSTEEKT	287
ushuman	QNPKLFSSTFV	DTVKIAEDKN	WPTGR...AKS	NTSLKKGDTN	SNAKAVATDL	TK.LTAEETK	296
usbear	NEHKLFLSKFV	SLTGVNENEN	WPTGSGATSS	NTTVVKGSTN	SNAKAVATDL	V.ALNRDEKT	294
swdog	TEKGGLLNKFV	SLTRVSDGTH	WPTGQGG.AS	GAGGQKQQLN	SNAKAMAKDL	VEQLTPEEKT	296
norsheep	SSPQYLKEFV	EKTLKDGSKN	WPTSTTGGAY.	...KTN	DNAEAVAGDL	TK.LTTEEKT	289
uswrat	IVAGLLAKTI	EGGEVVEIRA	VSSTSVMVNA	CYDLLSEGLG	VVPYACVGLG	GNFVGVVDGH	347
ushuman	IVAGLLAKTI	EGGEVVEIRA	VSSTSVMVNA	CYDLLSEGLG	VVPYACVGLG	GNFVGVVDGH	356
usbear	IVAGLLAKTI	EGGEVVEIRA	VSSTSVMVNA	CYDLLSEGLG	VVPYACVGLG	GNFVGVVDGH	354
swdog	IVAGLLAKTI	EGGEVVEIRA	VSSTSVMVNA	CYDLLSEGLG	VVPYACVGLG	GNFVGVVDGH	356
norsheep	IVAGLLAKTI	EGGEVVEIRA	VSSTSVMVNA	CYDLLSEGLG	VVPYACVGLG	GNFVGVVDGH	349
uswrat	ITPKLAYRLK	AGLSYQLSPE	ISAFAGGFYH	RVVGDGVYDD	LPAQRLVDDT	SPAGRTKDTA	407
ushuman	ITPKLAYRLK	AGLSYQLSPE	ISAFAGGFYH	RVVGDGVYDD	LPAQRLVDDT	SPAGRTKDTA	416
usbear	ITPKLAYRLK	AGLSYQLSPE	ISAFAGGFYH	RVVGDGVYDD	LPAQRLVDDT	SPAGRTKDTA	414
swdog	ITPKLAYRLK	AGLSYQLSPE	ISAFAGGFYH	RVVGDGVYDD	LPAQRLVDDT	SPAGRTKDTA	416
norsheep	ITPKLAYRLK	AGLSYQLSPE	ISAFAGGFYH	RVVGDGVYDD	LPAQRLVDDT	SPAGRTKDTA	409
uswrat	IANFPMAVVG	GEFGVRFAP*					426
ushuman	IANFMSMAVVG	GEFGVRFAP*					435
usbear	IANFMSMAVVG	GEFGVRFAP*					433
swdog	IANFMSMAVVG	GEFGVRFAP*					435
norsheep	IANFMSMAVVG	GEFGVRFAP*					428

FIG. 2. Variation in the amino acid sequences of P44ESup1 (A) or MSP2 (P44) (B) between U.S. and European strains of *A. phagocytophilum* and strains infecting different mammalian host species (ushuman, U.S. human; uswrat, U.S. wood rat, *Neotoma fuscipes*; usbear, U.S. bear, *Ursus americanus*; swdog, Swedish dog, *Canis familiaris*; and norsheep, Norwegian domestic sheep, *Ovis aries*). The sequence of a strain infecting a human patient in the United States is from GenBank AY164493; all other sequences are from this study.

TABLE 1. Number of shared expression site variants with HZ strain genomic loci

Origin of <i>A. phagocytophilum</i>	No. of variants shared/total ^a
Cultured HGE2 strain	16/17
Cultured NY-18 strain	5/5
Cultured Webster strain	3/3
Humans	26/26
Bears	3/4
Wood rats	5/19
Norwegian sheep	0/23
Swedish dogs	1/19

^a The number of shared expression site variants was calculated at $\geq 90\%$ amino acid sequence identity.

strains infecting different Swedish dogs, and 3 were among strains infecting different Norwegian sheep. Therefore, some similar variants were observed in populations derived from humans, wood rats (a potential reservoir species; for details, see references 17 and 18), and a bear in the United States and also in organisms derived from humans and grown in culture. However, there was no sharing at $\geq 90\%$ identity between the European and U.S. strains, or even between the strains from Norway and Sweden. MSP2 (P44) paralogs in the databases have not generally been identified as to their genomic origin (i.e., expression site versus pseudogene). However, we also compared the 116 expression site sequences from this study with all available paralogs in the UniProt/Swiss-Prot database (Batch BLAST; http://services.ncbi.nlm.nih.gov/cgi-bin/bb_search.cgi). Zero of 19 Swedish dog variants and 5 of 23 Norwegian sheep variants matched a database paralog at $\geq 90\%$ identity. Four of the five Norwegian sheep expression site variants matched paralogs previously identified in strains of *A. phagocytophilum* in sheep from the United Kingdom (11).

The complete genome sequence of a U.S. (human-derived) strain, HZ, of *A. phagocytophilum* has recently been published (14). A total of 113 *p44* loci were annotated, but some of these are truncated or are short 5' or 3' gene fragments. Within these 113 loci, we identified 94 that had at least one of the conserved regions flanking the central hypervariable region that previous data (28) suggest is required for gene conversion. The presence of at least one of these flanking conserved regions also enables a reasonable alignment to be made. A MatGAT comparison was then conducted that included these 94 genomic *p44* loci identified in the complete *A. phagocytophilum* genome sequence as well as the 116 expression site variants from the United States and Europe (see Fig. S1 in the supplemental material). We identified those pairwise alignments having 100 and $\geq 90\%$ identity, in order not to miss otherwise good matches that differed by only a few amino acids. With the exception of wood rat variants, we could match nearly all U.S.-origin *A. phagocytophilum* expression site variants with a locus in the sequenced HZ strain (Table 1). There were many 100% identities between the different genomic copies and U.S. human or human-derived culture variants (see Table S1 in the supplemental material). In addition, some *A. phagocytophilum* variants infecting wood rats or bears in the United States matched at $\geq 90\%$ identity with known genomic copies of the HZ strain. However, this was not the case with the European variants. No alignments were found having

$\geq 90\%$ identity between the genomic loci and the variants identified in *A. phagocytophilum* infections of Norwegian sheep. There was one match at $\geq 90\%$ identity between a Swedish dog variant (D1v1) and a (U.S.) genomic copy (aph_1128). An examination of this alignment (see Fig. S1 in the supplemental material) reveals a very similar hypervariable region that likely results from limited evolutionary divergence of the same donor pseudogene copy. Despite the general dissimilarity between genomic copies of the U.S. strain and European expression site variants, the overall structure of the central hypervariable region was maintained in the European strains, in terms of the known signature conserved residues (29). This suggests that while an analogous variation mechanism is used in the European strains, there is diversity in the genomic donor pseudogene repertoire used for gene conversion of the expression site.

A second potentially contributing factor to MSP2 (P44) diversity was apparent from the analysis of *A. phagocytophilum* populations found in the wood rats sampled at one location in Hoopa Valley, California. The alignment of *msp2* (*p44*) expression site sequences from variants found at similar times at this field site revealed the common occurrence of variants with conserved sequence blocks throughout the hypervariable region (Fig. 3). These wood rat variants had sequence identities ranging between 41 and 89%. For example, wood rat 3 contained a variant (WR3v1) that was identical to a variant in wood rat 5 (WR5v3) except for short segments in the center of the hypervariable region and close to the 5' conserved sequence. Wood rat 3 also contained a variant (WR3v2) sharing shorter identical segments with WR4v4; similarly, shared blocks of sequence were present in the three variants WR1v1, WR2v2, and WR6v1. This mosaic structure has been commonly observed throughout the hypervariable region in *A. marginale* infections (3) but not previously in *A. phagocytophilum*.

DISCUSSION

Previous data on MSP2 (P44) expression, structure and diversity have been obtained primarily from strains of *A. phagocytophilum* derived from human infections. We have extended this analysis to include strains from a wide range of animal species that are known to be naturally infected, including infections of animals from the United States and Europe, domestic ruminants, and potential host reservoir species. The data reveal similar expression site structures for MSP2 (P44) in all strains. Conserved structural elements include the upstream promoter and downstream truncated *recA* gene. The promoter regions in strains derived from infections of dogs and sheep in Europe are identical through the -35 and -10 regions and also include the 12-base palindrome that encompasses the transcription start site (2). In both of the European strains examined, the only change in this region was a single base insertion between the -10 region and the palindrome. Analogous to human strains from the United States, the expression site includes the coding sequence for MSP2 (P44) and an upstream gene, *P44ESup1*. Although recognizable as most homologous to P44ESup1 from U.S. human strains, the P44ESup1 from Norwegian sheep had diverged considerably and contained many amino acid substitutions throughout the molecule. Nucleotide regions surrounding *msp2* (*p44*) in the

expression site had also diverged, rendering the previous primers for PCR amplification of the expression site (designed from the sequence of U.S. strains; for details, see reference 4) unusable. Nevertheless, MSP2 (P44) itself was well conserved and had N and C termini nearly identical to those of U.S. strains and also contained the central hypervariable region previously implicated in the generation of antigenic diversity through gene conversion mechanisms (28). Another conserved feature was the presence of multiple different variant types in each *A. phagocytophilum* population, whether derived from infections of humans, bears, or wood rats in the United States or sheep and dogs in Europe. The variant types also differed at different sampling time points in persistent infections of sheep. This suggests that similar mechanisms of persistence through gene conversion of this expression site are available to the organisms infecting these diverse animal species.

Out of 6,670 pairwise alignments of expression site variants from the United States and Europe, only about 1% had significant identity (defined here as $\geq 90\%$ identity in the hypervariable region). None of the pairwise alignments with $\geq 90\%$ identity were between *A. phagocytophilum* isolates of U.S. and European origin or between isolates of Swedish and Norwegian origin. This suggests that the repertoire of outer membrane protein variants that can potentially be expressed globally by *A. phagocytophilum* is very large. The great flexibility of this organism, with respect to its genome structure and its ability to infect diverse vertebrate and invertebrate hosts, is unusual for an organism of only 1.47 Mb total genome size.

In an analysis of 116 total expression site variants encoding MSP2 (P44), we identified matches with $\geq 90\%$ identity to 36 genomic loci of the sequenced U.S. HZ strain of *A. phagocytophilum* and matches of 100% identity to 18 loci. These loci have been annotated as two different gene classes, either full-length or silent/reserve genes (14). The presence of sequence from both classes in the expression site supports the previous suggestion that either class can donate sequence to the expression site to generate outer membrane protein diversity (14). The expression site variants that could be matched to genomic loci were uniformly derived from *A. phagocytophilum* of U.S. origin, with the single exception of one variant of Swedish origin. This suggests marked diversity in the strains of *A. phagocytophilum* isolated in Europe compared to that in North American strains. It also supports previous microarray data showing that all but four *msp2* (p44) genomic copies were present in three U.S. strains from different geographic locations (14). However, these analyses are still limited to relatively few populations of *A. phagocytophilum* and have not yet included, for example, populations circulating in white-tailed deer in the United States that are thought not to be associated with human infection (30).

Only 5/19 wood rat variants could be matched to known genome loci, in contrast to the matching of nearly all variants isolated directly from humans or bears or grown in culture. The human variants were derived from acute, clinically symptomatic cases. Although the times of infection of the wood rats are not known, these animals undergo persistent infections with *A. phagocytophilum* and have been implicated as a potential reservoir species (17). Therefore, it is likely that the wood rat samples contain organisms more representative of long-term infections. The shared blocks of sequence found through-

out the hypervariable region in variants derived from different wood rats suggest that segmental gene conversion may introduce more-complex mosaics into the expression site, as has been shown in *A. marginale* (9). If it is similar to *A. marginale*, the basic pseudogene repertoire may be expressed earlier in infections without extensive recombination among hypervariable region sequences (21). Therefore, resolving this question definitively will require the analysis of long-term infections in reservoir hosts experimentally infected with a genome-sequenced strain of *A. phagocytophilum*.

Unlike wood rats, sheep, and dogs, where persistent infections have been clearly demonstrated (15, 18, 39), it is unclear if persistent infections are a feature of the disease in humans. However, this possibility should be considered in light of the extensive structural diversity demonstrated here, the long-term adverse health outcomes of *A. phagocytophilum* infection (36), and the high rate of serological positivity (15 to 36%) of people in areas of endemicity (13). Another important question is whether antibiotic treatment, as currently prescribed, completely sterilizes the infection.

Despite the expression site diversity seen in this study, it is clearly not limitless even within the central hypervariable region of MSP2 (P44). In addition to the conserved "signature" residues KVC, C, F, and NWPT, there are conservative replacements elsewhere with respect to charge or size that presumably form the necessary framework for correct folding of the protein on the surface of the organism. The different hypervariable regions could have additional functions beyond immune evasion, such as in binding to and invading different cell types. There is evidence that antibodies to MSP2 (P44) and recombinant MSP2 (P44) antagonize adhesion of *A. phagocytophilum* to granulocytes (13, 35). The binding of bacteria to particular cell types may depend on the expression of particular MSP2 (P44) variants. There are now many orthologs and paralogs of MSP2 (P44) known in different *Anaplasma* and *Ehrlichia* species that form Pfam01617. No three-dimensional structure is known for any of these; therefore, an important priority for future research should be to crystallize these outer membrane proteins and obtain complete structures. This would allow a better interpretation of the biological significance of the most variable and more conserved framework residues. The data presented here should allow the correlation of sequence changes with structure and with effects on immune evasion and host cell binding.

In summary, *A. phagocytophilum* organisms infecting diverse animal species and humans in the United States and Europe appear to possess and utilize similar genomic expression loci for generating outer membrane protein diversity. Although there has been some divergence in this locus between U.S. and European strains, fundamental structural features are conserved. There is significant global diversity in MSP2 (P44) generated through recombination into this locus. It may represent an ongoing adaptation to cell invasion and survival in different hosts as well as an obstacle to the control of this persistent infection.

ACKNOWLEDGMENTS

This investigation received support from NIH grant AI45580.

We thank Robert Massung of the Centers for Disease Control, Atlanta, GA, for assistance with the provision of DNA samples from

European sources of *A. phagocytophilum* and Ulrike G. Munderloh and Susan J. Wong for provision of samples from cultures of the HGE2 strain and from infected patients.

REFERENCES

1. Agüero, R. M., H. W. Horowitz, G. P. Wormser, D. F. McKenna, J. Nowakowski, J. Munoz, and J. S. Dumler. 1996. Human granulocytic ehrlichiosis: a case series from a medical center in New York State. *Ann. Intern. Med.* **125**:904–908.
2. Barbet, A. F., J. T. Agnes, A. L. Moreland, A. M. Lundgren, A. R. Alleman, S. M. Noh, K. A. Brayton, U. G. Munderloh, and G. H. Palmer. 2005. Identification of functional promoters in the msp2 expression loci of *Anaplasma marginale* and *Anaplasma phagocytophilum*. *Gene* **353**:89–97.
3. Barbet, A. F., A. M. Lundgren, J. Yi, F. R. Rurangirwa, and G. H. Palmer. 2000. Antigenic variation of *Anaplasma marginale* by expression of MSP2 mosaics. *Infect. Immun.* **68**:6133–6138.
4. Barbet, A. F., P. F. Meeus, M. Belanger, M. V. Bowie, J. Yi, A. M. Lundgren, A. R. Alleman, S. J. Wong, F. K. Chu, U. G. Munderloh, and S. D. Jauron. 2003. Expression of multiple outer membrane protein sequence variants from a single genomic locus of *Anaplasma phagocytophilum*. *Infect. Immun.* **71**:1706–1718.
5. Barbet, A. F., J. Yi, A. M. Lundgren, B. R. McEwen, E. F. Blouin, and K. M. Kocan. 2001. Antigenic variation of *Anaplasma marginale*: major surface protein 2 diversity during cyclic transmission between ticks and cattle. *Infect. Immun.* **69**:3057–3066.
6. Brayton, K. A., L. S. Kappmeyer, D. R. Herndon, M. J. Dark, D. L. Tibbals, G. R. Alleman, T. C. McGuire, and D. P. Knowles, Jr. 2005. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc. Natl. Acad. Sci. USA* **102**:844–849.
7. Brayton, K. A., D. P. Knowles, T. C. McGuire, and G. H. Palmer. 2001. Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens. *Proc. Natl. Acad. Sci. USA* **98**:4130–4135.
8. Brayton, K. A., P. F. Meeus, A. F. Barbet, and G. H. Palmer. 2003. Simultaneous variation of the immunodominant outer membrane proteins, MSP2 and MSP3, during *Anaplasma marginale* persistence in vivo. *Infect. Immun.* **71**:6627–6632.
9. Brayton, K. A., G. H. Palmer, A. M. Lundgren, J. Yi, and A. F. Barbet. 2002. Antigenic variation of *Anaplasma marginale* msp2 occurs by combinatorial gene conversion. *Mol. Microbiol.* **43**:1151–1159.
10. Campanella, J. J., L. Bitincka, and J. Smalley. 2003. MatGAT: an application that generates similarity/identity matrices using protein or DNA sequences. *BMC Bioinformatics* **4**:29–32.
11. Casey, A. N., R. J. Birtles, A. D. Radford, K. J. Bown, N. P. French, Z. Woldehiwet, and N. H. Ogden. 2004. Groupings of highly similar major surface protein (p44)-encoding paralogues: a potential index of genetic diversity amongst isolates of *Anaplasma phagocytophilum*. *Microbiology* **150**:727–734.
12. Dumler, J. S. 2005. *Anaplasma* and *ehrlichia* infection. *Ann. N. Y. Acad. Sci.* **1063**:361–373.
13. Dumler, J. S., K. S. Choi, J. C. Garcia-Garcia, N. S. Barat, D. G. Scarpio, J. W. Garyu, D. J. Grab, and J. S. Bakken. 2005. Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*. *Emerg. Infect. Dis.* **11**:1828–1834.
14. Dunning Hotopp, J. C., M. Lin, R. Madupu, J. Crabtree, S. V. Angiuoli, J. Eisen, R. Seshadri, Q. Ren, M. Wu, T. R. Utterback, S. Smith, M. Lewis, H. Khouri, C. Zhang, H. Niu, Q. Lin, N. Ohashi, N. Zhi, W. Nelson, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, J. Sundaram, S. C. Daugherty, T. Davidsen, A. S. Durkin, M. Gwinn, D. H. Haft, J. D. Selengut, S. A. Sullivan, N. Zafar, L. Zhou, F. Benahmed, H. Forberger, R. Halpin, S. Mulligan, J. Robinson, O. White, Y. Rikihisa, and H. Tettelin. 2006. Comparative genomics of emerging human ehrlichiosis agents. *PLOS Genet.* **2**:208–223.
15. Egenvall, A., I. Lilliehook, A. Bjoersdorff, E. O. Engvall, E. Karlstam, K. Artursson, M. Heldtander, and A. Gunnarsson. 2000. Detection of granulocytic *Ehrlichia* species DNA by PCR in persistently infected dogs. *Vet. Rec.* **146**:186–190.
16. Eid, G., D. M. French, A. M. Lundgren, A. F. Barbet, T. F. McElwain, and G. H. Palmer. 1996. Expression of major surface protein 2 antigenic variants during acute *Anaplasma marginale* rickettsemia. *Infect. Immun.* **64**:836–841.
17. Foley, J. E., P. Foley, R. N. Brown, R. S. Lane, J. S. Dumlers, and J. E. Madigan. 2004. Ecology of *Anaplasma phagocytophilum* and *Borrelia burgdorferi* in the western United States. *J. Vector Ecol.* **29**:41–50.
18. Foley, J. E., V. Kramer, and D. Weber. 2002. Experimental infection of dusky-footed wood rats (*Neotoma fuscipes*) with *Ehrlichia phagocytophila sensu lato*. *J. Wildl. Dis.* **38**:194–198.
19. French, D. M., W. C. Brown, and G. H. Palmer. 1999. Emergence of *Anaplasma marginale* antigenic variants during persistent rickettsemia. *Infect. Immun.* **67**:5834–5840.
20. French, D. M., T. F. McElwain, T. C. McGuire, and G. H. Palmer. 1998. Expression of *Anaplasma marginale* major surface protein 2 variants during persistent cyclic rickettsemia. *Infect. Immun.* **66**:1200–1207. (Erratum, **66**:2400.)

21. Futse, J. E., K. A. Brayton, D. P. Knowles, Jr., and G. H. Palmer. 2005. Structural basis for segmental gene conversion in generation of *Anaplasma marginale* outer membrane protein variants. *Mol. Microbiol.* **57**:212–221.
22. Goodman, J. L., C. Nelson, B. Vitale, J. E. Madigan, J. S. Dumler, T. J. Kurtti, and U. G. Munderloh. 1996. Direct cultivation of the causative agent of human granulocytic ehrlichiosis. *N. Engl. J. Med.* **334**:209–215. (Erratum, 335:361.)
23. Goodstadt, L., and C. P. Ponting. 2001. CHROMA: consensus-based colouring of multiple alignments for publication. *Bioinformatics* **17**:845–846.
24. Horowitz, H. W., T. C. Hsieh, M. E. Aguero-Rosenfeld, F. Kalantarpour, I. Chowdhury, G. P. Wormser, and J. M. Wu. 2001. Antimicrobial susceptibility of *Ehrlichia phagocytophila*. *Antimicrob. Agents Chemother.* **45**:786–788.
25. Ijdo, J. W., W. Sun, Y. Zhang, L. A. Magnarelli, and E. Fikrig. 1998. Cloning of the gene encoding the 44-kilodalton antigen of the agent of human granulocytic ehrlichiosis and characterization of the humoral response. *Infect. Immun.* **66**:3264–3269.
26. Lin, Q., and Y. Rikihisa. 2005. Establishment of cloned *Anaplasma phagocytophilum* and analysis of p44 gene conversion within an infected horse and infected SCID mice. *Infect. Immun.* **73**:5106–5114.
27. Lin, Q., Y. Rikihisa, N. Ohashi, and N. Zhi. 2003. Mechanisms of variable p44 expression by *Anaplasma phagocytophilum*. *Infect. Immun.* **71**:5650–5661.
28. Lin, Q., C. Zhang, and Y. Rikihisa. 2006. Analysis of involvement of the RecF pathway in p44 recombination in *Anaplasma phagocytophilum* and in *Escherichia coli* by using a plasmid carrying the p44 expression and p44 donor loci. *Infect. Immun.* **74**:2052–2062.
29. Lin, Q., N. Zhi, N. Ohashi, H. W. Horowitz, M. E. Aguero-Rosenfeld, J. Raffalli, G. P. Wormser, and Y. Rikihisa. 2002. Analysis of sequences and loci of p44 homologs expressed by *Anaplasma phagocytophila* in acutely infected patients. *J. Clin. Microbiol.* **40**:2981–2988.
30. Massung, R. F., T. N. Mather, and M. L. Levin. 2006. Reservoir competency of goats for the Ap-variant 1 strain of *Anaplasma phagocytophilum*. *Infect. Immun.* **74**:1373–1375.
31. Meeus, P. F., K. A. Brayton, G. H. Palmer, and A. F. Barbet. 2002. Conservation of a gene conversion mechanism in two distantly related paralogs of *Anaplasma marginale*. *Mol. Microbiol.* **47**:633–643.
32. Morgenstern, B. 1999. DIALIGN 2: improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* **15**:211–218.
33. Murphy, C. L., J. R. Storey, J. Recchia, L. A. Doros-Richert, C. Gingrich-Baker, K. Munroe, J. S. Bakken, R. T. Coughlin, and G. A. Beltz. 1998. Major antigenic proteins of the agent of human granulocytic ehrlichiosis are encoded by members of a multigene family. *Infect. Immun.* **66**:3711–3718.
34. Ogden, N. H., Z. Woldehiwet, and C. A. Hart. 1998. Granulocytic ehrlichiosis: an emerging or rediscovered tick-borne disease? *J. Med. Microbiol.* **47**:475–482.
35. Park, J., K. S. Choi, and J. S. Dumler. 2003. Major surface protein 2 of *Anaplasma phagocytophilum* facilitates adherence to granulocytes. *Infect. Immun.* **71**:4018–4025.
36. Ramsey, A. H., E. A. Belongia, C. M. Gale, and J. P. Davis. 2002. Outcomes of treated human granulocytic ehrlichiosis cases. *Emerg. Infect. Dis.* **8**:398–401.
37. Rodriguez, J. L., G. H. Palmer, D. P. Knowles, Jr., and K. A. Brayton. 2005. Distinctly different msp2 pseudogene repertoires in *Anaplasma marginale* strains that are capable of superinfection. *Gene* **361**:127–132.
38. Rutherford, K., J. Parkhill, J. Crook, T. Horsnell, P. Rice, M. A. Rajandream, and B. Barrell. 2000. Artemis: sequence visualization and annotation. *Bioinformatics* **16**:944–945.
39. Stuenkel, S., A. N. Casey, Z. Woldehiwet, N. P. French, and N. H. Ogden. 2006. Detection by the polymerase chain reaction of *Anaplasma phagocytophilum* in tissues of persistently infected sheep. *J. Comp. Pathol.* **134**:101–104.
40. Zhi, N., N. Ohashi, Y. Rikihisa, H. W. Horowitz, G. P. Wormser, and K. Hechemy. 1998. Cloning and expression of the 44-kilodalton major outer membrane protein gene of the human granulocytic ehrlichiosis agent and application of the recombinant protein to serodiagnosis. *J. Clin. Microbiol.* **36**:1666–1673.

Editor: W. A. Petri, Jr.