Sequence Analysis of p44 Homologs Expressed by *Anaplasma phagocytophilum* in Infected Ticks Feeding on Naive Hosts and in Mice Infected by Tick Attachment

Suleyman Felek,1 Sam Telford III,2 Richard C. Falco,3 and Yasuko Rikihisa1*

Department of Veterinary Biosciences, College of Veterinary Medicine, The Ohio State University, Columbus, Ohio 43210-1093;1
Division of Infectious Diseases, School of Veterinary Medicine, Tufts University, North Grafton, Massachusetts 01536;2
and Vector Ecology Laboratory, Louis Calder Center, Fordham University, Armonk, New York 105043

Received 17 July 2003/Returned for modification 22 September 2003/Accepted 11 November 2003

The 44-kDa immunodominant outer membrane proteins (P44 proteins) of *Anaplasma phagocytophilum* are encoded by the p44 polymorphic multigene family. The present study examined p44 expression and analyzed the cDNA sequences of various p44 transcripts from the spleens and blood of mice infected by the bites of ticks infected with the *A. phagocytophilum* NTN-1 strain or of naturally infected nymphal ticks and in the salivary glands and midgut tissues of these ticks. A total of 300 p44 cDNAs were subjected to sequence analysis. Of these, 40 distinct p44 species were found, and all of these had orthologs in the *A. phagocytophilum HZ* strain genome that shared 95 to 100% base sequence identity. The number of unique p44 species expressed in mouse blood was greater than that for mouse spleens. Higher numbers of different p44 transcripts were also expressed in the salivary glands of ticks than in the midgut tissues. Variations in the sequences of the same p44 cDNA species within a single *A. phagocytophilum* strain and among different strains were concentrated in the conserved regions flanking the central hypervariable region of p44 genes. No mosaic sequences derived from two or more p44 species were found within the p44 hypervariable region. The conservation of the hypervariable region of each p44 cDNA species of *A. phagocytophilum* in naturally infected ticks and in different geographic isolates suggests that each *A. phagocytophilum* genome carries a set of p44 paralogs to be expressed. Thus, a large but restricted repertoire of p44 hypervariable sequences exists in *A. phagocytophilum* strains in the Northeastern United States.

Human granulocytic ehrlichiosis (HGE) is a tick-borne zoonosis that was first reported in the United States in 1994. This disease has been documented in North America and in Europe and was designated in 1998 as a nationally reportable disease in the United States (22). *Anaplasma phagocytophilum*, the etiologic agent of HGE, is a gram-negative, obligatory intracellular bacterium that primarily infects neutrophils (10). In the Eastern and Midwestern United States, *A. phagocytophilum* is transmitted by *Ixodes scapularis* ticks (8, 33). *Ixodes pacificus* and *Ixodes ricinus* have been identified as vectors in California (30) and Europe (26), respectively. The major reservoir of *A. phagocytophilum* in the Northeastern and Midwestern United States is the white-footed mouse (*Peromyscus leucopus*), which may be infected either transiently or persistently (29, 34). Although various wild and domestic animal species have been reported to be infected with *A. phagocytophilum* (10), to date the 16S rRNA gene sequence corresponding to *A. phagocytophilum* that infects humans (the HGE agent) has been identified only in white-footed mice (29, 34, 37), horses (16a, 27, 36), and dogs (16a, 28).

The incubation period of HGE is 1 to 2 weeks after the tick bite occurs (2). The most common symptoms and signs of the disease include fever, myalgia, rigors, headache, and malaise (2, 11). Most patients also exhibit leukopenia, thrombocytopenia, and elevated levels of hepatic transaminases (2, 11). The fatality rate is approximately 2 to 5% (22), and deaths have occurred primarily in patients that acquire opportunistic infections after contracting HGE (11).

Outer membrane proteins of 44 to 49 kDa have been shown to act as the major antigens recognized by human sera (1, 15, 35, 38–40). The 44-kDa major outer membrane proteins (P44 proteins) of *A. phagocytophilum* are encoded by the p44 polymorphic multigene family. The p44 genes are dispersed in the genome (38, 40). Based on the preliminary *A. phagocytophilum HZ* genome sequence (http://www.tigr.org), the total number of p44 paralogs is >80. p44 genes consist of a central hypervariable region, of approximately 280 bp, which is flanked by conserved sequences. Many, but not all, p44 genes lack start codons (38, 40, 41). Interleukin-1β, tumor necrosis factor-α, and interleukin-6 are generated when human peripheral blood leukocytes are exposed to recombinant P44-1 protein (18). This release of proinflammatory cytokines may explain the clinical signs and hematological abnormalities associated with HGE and suggests that P44 proteins play a role in the pathogenesis of this disease (18). Members of our laboratory previously reported that passive immunization of mice with monoclonal antibodies against P44 proteins partially protects them against *A. phagocytophilum* infection (17). Thus, P44 proteins may represent potential vaccine candidates for HGE.

Expression studies showed that p44 genes are differentially expressed in cultured cells and ticks and in the blood of mice, horses, and human patients (4, 14, 20, 40, 41). p44 is homologous to *msp2* of *Anaplasma marginale*, which is a bovine in-
traerythrocytic agent (24, 38, 40). It has been reported that the segmental gene conversion of multiple msp2 genes at a single expression locus in A. marginale allows unlimited variation of the expressed msp2 chimera or mosaic (3, 6, 7). Recently, Barbet et al. proposed that the same mechanism controls the expression of variable p44 genes in A. phagocytophilum (4). If this is true, the hypervariable region sequences of p44 in the expression locus will be chimeras or mosaics of two or more donor p44 genes elsewhere in the same genome. However, this contrasts with previous data which suggested that the hypervariable region sequences in several p44 paralogs cloned from different genomic loci are identical in the corresponding regions of p44 CDNA (40, 41). For the present study, we investigated p44 expression in the blood and spleens of strain DBA/2 mice infected by attaching I. scapularis ticks to simulate natural infection and in salivary glands and midgut tissues of infected ticks feeding on naive hosts (transmission-fed ticks). These ticks were infected either naturally or experimentally with the NTN-1 strain of A. phagocytophilum. The NTN-1 strain was maintained by alternating tick-mouse passages (33). P44 CDNA sequences from field ticks and the NTN-1 strain were compared with p44 DNA sequences of the HZ strain. In addition, individual cDNA sequences of the same p44 species, both within the same strain and among different strains, were compared to identify sequence differences and variable regions.

MATERIALS AND METHODS

Tick attachment. One hundred I. scapularis nymphs were collected in Westchester County, N.Y., by drag sampling (12). Ticks were kept in an incubator at 18°C under 95 to 100% relative humidity with a 12-h photoperiod for more than 10 days, until they were attached to male DBA/2 mice (5 to 6 weeks old) (Harlan Sprague-Dawley, Indianapolis, Ind.).

The A. phagocytophilum NTN-1 strain used was in approximately the 14th passage (tick-mouse cycle). The original isolate was obtained by inoculating a C3H mouse with Feulgen-positive salivary glands from ticks collected from the yard of the index patient in Nantucket, Mass. (33). This mouse served as the host for lab-reared ticks. The strain is maintained by allowing infected ticks to feed on SCID mice infected with the NTN-1 strain of A. phagocytophilum (24, 38, 40). It has been reported that the hypervariable region sequences of p44 (4). If this is true, the hypervariable region sequences of p44 in the expression locus will be chimeras or mosaics of two or more donor p44 genes elsewhere in the same genome. However, this contrasts with previous data which suggested that the hypervariable region sequences in several p44 paralogs cloned from different genomic loci are identical in the corresponding regions of p44 CDNA (40, 41). For the present study, we investigated p44 expression in the blood and spleens of strain DBA/2 mice infected by attaching I. scapularis ticks to simulate natural infection and in salivary glands and midgut tissues of infected ticks feeding on naive hosts (transmission-fed ticks). These ticks were infected either naturally or experimentally with the NTN-1 strain of A. phagocytophilum. The NTN-1 strain was maintained by alternating tick-mouse passages (33). P44 CDNA sequences from field ticks and the NTN-1 strain were compared with p44 DNA sequences of the HZ strain. In addition, individual cDNA sequences of the same p44 species, both within the same strain and among different strains, were compared to identify sequence differences and variable regions.

Tick attachment. One hundred I. scapularis nymphs were collected in Westchester County, N.Y., by drag sampling (12). Ticks were kept in an incubator at 18°C under 95 to 100% relative humidity with a 12-h photoperiod for more than 10 days, until they were attached to male DBA/2 mice (5 to 6 weeks old) (Harlan Sprague-Dawley, Indianapolis, Ind.).

The A. phagocytophilum NTN-1 strain used was in approximately the 14th passage (tick-mouse cycle). The original isolate was obtained by inoculating a C3H mouse with Feulgen-positive salivary glands from ticks collected from the yard of the index patient in Nantucket, Mass. (33). This mouse served as the host for lab-reared ticks. The strain is maintained by allowing infected ticks to feed on SCID mice infected with the NTN-1 strain of A. phagocytophilum (24, 38, 40). It has been reported that the hypervariable region sequences of p44 (4). If this is true, the hypervariable region sequences of p44 in the expression locus will be chimeras or mosaics of two or more donor p44 genes elsewhere in the same genome. However, this contrasts with previous data which suggested that the hypervariable region sequences in several p44 paralogs cloned from different genomic loci are identical in the corresponding regions of p44 CDNA (40, 41). For the present study, we investigated p44 expression in the blood and spleens of strain DBA/2 mice infected by attaching I. scapularis ticks to simulate natural infection and in salivary glands and midgut tissues of infected ticks feeding on naive hosts (transmission-fed ticks). These ticks were infected either naturally or experimentally with the NTN-1 strain of A. phagocytophilum. The NTN-1 strain was maintained by alternating tick-mouse passages (33). P44 CDNA sequences from field ticks and the NTN-1 strain were compared with p44 DNA sequences of the HZ strain. In addition, individual cDNA sequences of the same p44 species, both within the same strain and among different strains, were compared to identify sequence differences and variable regions.

Tick attachment. One hundred I. scapularis nymphs were collected in Westchester County, N.Y., by drag sampling (12). Ticks were kept in an incubator at 18°C under 95 to 100% relative humidity with a 12-h photoperiod for more than 10 days, until they were attached to male DBA/2 mice (5 to 6 weeks old) (Harlan Sprague-Dawley, Indianapolis, Ind.).
(NTN-1 strain) and were subjected to nested PCR with p44 primers. One of the 10 naturally infected ticks and 1 of the 5 experimentally infected ticks were found to be positive. Of DNAs extracted from blood samples derived from 26 individual mice, specimens from 1 of 16 mice hosting naturally infected ticks (one tick per mouse) and from 9 of 10 mice hosting experimentally infected ticks (five ticks per mouse) were found to be positive for *A. phagocytophilum* infection by nested PCR.

**Differential expression of p44 in various tissues.** For identification of the specific p44 transcripts expressed in mice and ticks, samples were subjected to p44 RT-PCR. The resulting p44 cDNA PCR products were sequenced on both strands. Twenty-five cDNA clones from blood samples and 25 cDNA clones from spleen samples of each of the three infected mice (one with the Westchester strain and two with the NTN-1 strain) were sequenced. cDNA clones derived from salivary gland and midgut samples from one Westchester strain-infected tick or two pools of five ticks infected with the NTN-1 strain removed from each of these three mice were also analyzed, such that a total of 300 p44 cDNA clones were sequenced. Overall, a total of 40 different species of p44 were found to be transcribed in all specimens, with 27 different p44 species found in all blood samples alone (Fig. 1). Higher numbers of p44 species were found in mouse blood (13.3 ± 3.1; median, 14; *N* = 3) than in mouse spleens (6.3 ± 2.1; median, 7; *N* = 3) (*P* = 0.014). Higher numbers of p44 species were found in tick salivary glands (12.6 ± 1.2; median, 12; *N* = 3) than in tick midgut specimens (7.3 ± 3.6; median, 7; *N* = 3) (*P* = 0.014) (Table 1). Some sequences (p44-18 and p44-28) were found in three or

**TABLE 1. Numbers of different p44 cDNA species identified in A. phagocytophilum-infected mouse and tick tissues**

<table>
<thead>
<tr>
<th>Mouse no./tick ID</th>
<th>Mice</th>
<th>Ticks</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Spleen</td>
<td>Blood</td>
</tr>
<tr>
<td>4/Westchester</td>
<td>4</td>
<td>16</td>
</tr>
<tr>
<td>26/NTN-1</td>
<td>8</td>
<td>14</td>
</tr>
<tr>
<td>27/NTN-1</td>
<td>7</td>
<td>10</td>
</tr>
<tr>
<td>Average ± SD</td>
<td>6.3 ± 2.1&lt;sup&gt;a&lt;/sup&gt;</td>
<td>13.3 ± 3.1</td>
</tr>
</tbody>
</table>

<sup>a</sup> *P* = 0.014 (when compared to the number of p44 species expressed in the blood).

<sup>b</sup> *P* = 0.014 (when compared to the number of p44 species expressed in the midgut).

The Westchester tick was one naturally infected tick collected in Westchester County, N.Y. The NTN-1 ticks identification included five ticks experimentally infected with the NTN-1 strain of *A. phagocytophilum*.

**FIG. 1.** Numbers of each p44 transcript species found expressed by *A. phagocytophilum* NTN-1 or Westchester in various tissues taken from three infected mice and from infected ticks attached to the mice. (A) Mouse spleens; (B) mouse blood; (C) tick salivary glands; (D) tick midgut specimens. The vertical axis shows numbers of cDNA clones of each p44 species detected among a total of 300 cDNA clones sequenced. The horizontal axis shows the various p44 cDNA clone species.
more of the six mouse specimens analyzed, but not in any of the tick specimens. Other p44 species, such as p44-25, were found in more than three tick specimens, but not in mouse specimens, suggesting that some p44 species are preferentially expressed in either mammals or ticks (Fig. 1). We identified seven new p44 cDNA sequences (p44-51 to -57) that shared 61.3 to 85.2% nucleotide sequence identities with the closest known p44 cDNA or DNA sequences. These sequences have not yet been found to be expressed by the A. phagocytophilum HZ, NY-31, NY-36, or NY-37 strains. An alignment of the deduced amino acid sequences of all 40 expressed p44 species showed sequence identities ranging from 44 to 86%, as shown

FIG. 2. Alignment of representative deduced amino acid sequences of 40 p44 species found to be expressed in the present study. Aligned positions of identical amino acids are indicated by dashes. Gaps are indicated by dashed lines. The hypervariable region (40) is marked with a thick line (between two vertical arrows). Sequences P44-51 to P44-57 were identified for the first time in the present study.

61.3 to 85.2% nucleotide sequence identities with the closest known p44 cDNA or DNA sequences. These sequences have not yet been found to be expressed by the A. phagocytophilum HZ, NY-31, NY-36, or NY-37 strains. An alignment of the deduced amino acid sequences of all 40 expressed p44 species showed sequence identities ranging from 44 to 86%, as shown.
The same strain as well as among strains. These were unlikely to result from errors of Taq polymerase, since the same bases differed in both strands and often among several cDNA clones. When p44 cDNA sequences from the Westchester strain (from naturally infected ticks) and the NTN-1 strain (from experimentally infected ticks) were compared with the preliminary A. phagocytophilum HZ strain genome sequence, orthologs were found for all 40 cDNA species. The levels of identity between the HZ strain genome sequence and other cDNA base sequences were 95% for p44-54, 97% for p44-13, p44-20, and p44-56, and 98 to 100% for the remaining 36 different p44 sequences. To find regions of sequence variations within the same p44 cDNA species, we aligned the cDNA sequences within the same p44 species and strains and combined the data. Finally, the alignment data for each p44 species from a total of eight p44 species were combined. For this analysis, the eight p44 species that showed the most frequent sequence differences were selected, as shown in Fig. 4A. To find regions of sequence variations in the same p44 cDNA species among different strains, we aligned one representative majority p44 species cDNA sequence from each strain with the respective p44 ortholog DNA sequence for the HZ strain (35 NTN-1 strain cDNAs, 24 Westchester strain cDNAs, and 40 HZ strain DNA sequences, for a total of 99 sequences), as shown in Fig. 4B. The results of these cDNA sequence alignments showed that sequence differences among the same p44 cDNA species were found primarily in the 5'- and 3'-conserved regions flanking the central hypervariable regions of p44 genes (Fig. 4). The central hypervariable region base sequences (40) were highly conserved among the same p44 cDNA species across different strains as well as within each strain (Fig. 4). The DNA sequence variations were primarily found in the third base of the codon. Thus, these variations rarely changed the predicted amino acid sequences and did not alter the p44 cDNA species.

**DISCUSSION**

Antigenic changes in the bacterial surface play an important role in avoiding immunosurveillance and destruction by the host immune system. Polymorphic and immunodominant P44 proteins of A. phagocytophilum are the primary candidates for such a role. Antigenic variation is known to be used by several tick-borne bacterial pathogens to help them persist for long periods in the immunocompetent mammalian host, since tick feeding is seasonal. Furthermore, in regions where HGE is endemic, wild mammalian reservoirs frequently develop antibodies to the pathogen. Thus, the bacterial expression of diverse surface antigens likely improves the bacterium's odds of survival during tick transmission. In the present study, we identified several p44 species that were simultaneously expressed in tick and mouse tissues. Mouse blood and the tick salivary gland expressed a more diverse array of p44 species than the mouse spleen and the tick midgut. Blood and salivary glands are two tissues that may play a role in the transmission of ehrlichiae between mammalian and tick hosts, supporting the hypothesis that the diverse expression of various p44 species may be important for the transmission of A. phagocytophilum. Within a single tick, as many p44 species were transcribed as were found in pooled ticks. Differential expression of p44 genes diversifies the outer membrane protein properties of A.
phagocytophilum. Thus, this may also help the bacteria to adapt to and avoid innate immune responses in ticks. The diverse antigenic variation of p44 poses a challenge for the development of p44-based vaccines against A. phagocytophilum.

In the present study, we found 10 to 16 different p44 species that were expressed in blood samples from each of three mice 10 days after tick placement. Ijdo et al. reported that six different p44 species (of a total of 20 cDNA clones) were expressed in the blood specimens from one C3H/HeN mouse 10 days after intraperitoneal inoculation with SCID mouse blood infected with the NCH1 strain of A. phagocytophilum (14). In HGE patients, multiple p44 species were cotranscribed (4, 20).

With experimentally infected mammals, it was previously shown that two different p44 species of the HZ strain were transcribed in the blood of horses 8 days after inoculation or tick attachment, two different p44 species of the HZ strain were transcribed in mouse blood specimens 4 days after inoculation, and seven and two different p44 species, respectively, were expressed in the midguts and salivary glands of transmission-fed female tick samples (41). Jauron et al. showed that P44 antigens were expressed at low levels in ISE6 tick cells but that they were strongly expressed in human HL-60 cells (16). Thus, coexpression of diverse p44 species by A. phagocytophilum appears to be pervasive in various environments.

It was previously reported that p44-18 was the major transcript species in the A. phagocytophilum HZ strain found in horse and mouse blood at 8 and 4 days postinfection, respectively, but that this transcript was rarely detected in ticks (41). Also, expression of the p44-18 transcript was upregulated in A. phagocytophilum cultivated in HL-60 cells at 37°C compared to cells cultured at 24°C (41). In the present study, p44-18 transcripts were minor and were found only in mouse samples, not in tick samples. Thus, it appears that A. phagocytophilum strains or other factors determine the major p44 species expressed in mammals. Phylogenetic analysis of p44 cDNA sequences revealed that the p44 cDNAs are clustered into three
groups, the α, β, and γ groups. In the present study, p44-24 and p44-43, which are members of the γ group, were found to be expressed only in mouse specimens. Because only two p44 species were found for the γ group, we could not perform statistical analysis on this finding. Our sequence alignment showed that none of the six p44 species of the NCH1 strain expressed in transmission-fed ticks described by Ijdo et al. (14) belong to the γ group.

A. marginale is the most prevalent tick-transmitted bacterium in cattle and is closely related to A. phagocytophilum (19). Antigenic variation in the major surface protein 2 (MS2P) of A. marginale causes persistent infections of red blood cells in cattle. ms2 is a member of a multigene family. During infection in cattle, multiple MS2P variants occur every 6 to 8 weeks, and such antigenic variation allows the bacteria to escape from preexisting immune responses (13). Nine ms2 pseudogenes and a single complete gene have been reported in the St. Maries, Florida, and South Idaho strains of A. marginale (6). Recombination of multiple small segments of pseudogenes into the expression site by gene conversion generates a nearly infinite number of expressed ms2 sequences (3, 7). The p44 family of A. phagocytophilum shares a high degree of homology with the A. marginale ms2 family (24, 38, 40). Recently, Barbet et al. proposed that similar segmental conversion acts to control the differential expression of p44 genes by A. phagocytophilum (4). However, a recent study (21) suggested that differential p44 gene expression in A. phagocytophilum human isolates occurs through nonreciprocal conversion of the entire (nonsegmental) p44 hypervariable region, including parts of the conserved flanking region sequences, at the p44 expression locus to a sequence copied from one of the conserved p44 donor genomic loci. In the present study, we compared the cDNA sequences we obtained from the NTN-1 and Westchester tick strains with the A. phagocytophilum HZ strain genome sequence to investigate the possibility of combinatorial recombination such as that which occurs in A. marginale. We found that the HZ genome has p44 genes at each distinct locus that are orthologous to every p44 sequence (438 to 522 bp) expressed in the NTN-1 and Westchester tick strains. There were only between 0 and 5% differences among the same p44 species sequences in the HZ, NTN-1, and Westchester tick strains, and 36 of 40, or 90%, of the sequences, shared 98 to 100% identity. This indicates that despite the geographical difference between Nantucket Island, Mass. (NTN-1 strain), and Westchester County, N.Y., or between HGE patient isolates and bacteria isolated from field-collected ticks, the strain divergence for each p44 species was relatively low. Within a single strain or isolate, the identities of the same p44 species cDNA sequences from four tissues of mice and ticks were 97 to 100%. This further suggests that combinatorial recombination is undetectable in p44 genes of A. phagocytophilum and that the P44 antigenic repertoire at the hypervariable region is restricted.

The novel finding of the present study is that in each p44 cDNA species, conserved region base sequences flanking the central hypervariable (p44 species-specific) region were paradoxically more variable than those of the central hypervariable region within the same strain as well as across different strains. This result is consistent with our first hypothesis, which proposed nonsegmental gene conversion of the p44 hypervariable region (21). These findings also support our second hypothesis (21) that these divergent sequences in the flanking conserved regions result from heteroduplex formation between two nonidentical p44 sequences at both regions and from mismatch repair, as previously demonstrated for homologous genes of Escherichia coli (9).

The present results show that p44 expression patterns vary in different host tissues. Further studies are needed to understand the detailed mechanisms by which this diverse pattern of p44 expression occurs and to determine what factors facilitate this varied expression in different environments. These new findings will be useful for the eventual design of effective vaccines and therapeutic strategies against A. phagocytophilum.

ACKNOWLEDGMENTS

We thank O. Ozturk for statistical help and T. Boccia for assistance with field work.

This research was supported by grant R01 AI47407 from the National Institutes of Health. The genome of A. phagocytophilum was sequenced at The Institute for Genomic Research. The sequencing project was supported by National Institutes of Health grant R01 AI47885 to Y.R.

REFERENCES


