Establishment of Cloned *Anaplasma phagocytophilum* and Analysis of p44 Gene Conversion within an Infected Horse and Infected SCID Mice

Quan Lin and Yasuko Rikihisa*

Department of Veterinary Biosciences, The Ohio State University, 1925 Coffey Road, Columbus, Ohio 43210

Received 30 January 2005/Returned for modification 7 March 2005/Accepted 30 March 2005

Diverse p44 alleles at the p44 expression locus (p44Es) encoding surface-exposed major membrane proteins, P44s, of *Anaplasma phagocytophilum* were hypothesized to be garnered by recombination to enact antigenic variation. However, this hypothesis has not been proven so far, due to inability to clone this obligate intra-granulocytic rickettsia. To define the p44E recombination, we developed a novel method to clone *A. phagocytophilum*. This isogenic cloned population containing a defined p44E was used to infect a naïve horse and severe combined immunodeficiency (SCID) mice. During a 58-day infection period in the blood of the horse, p44E conversion was evident in a total of 11 new p44Es, 48% (115/242) of the sequenced p44E population. During a 50-day infection period in the blood of SCID mice, p44E conversion was manifested in a total of 13 new p44Es, 42% (192/460) of the p44E population. Thus, similar levels of p44E convertants were detected in either the presence or absence of an acquired immune system, suggesting that T- and B-cell immune pressure was not essential for recombination and/or selection of the p44E variants. Analysis of sequentially changed p44Es revealed that the entire central hypervariable region of donor p44 pseudogenes or of donor full-length p44s replaced the same region of the resident p44E as a cassette. Putative recombination points were detected within p44E conserved regions flanking the central hypervariable region by the TOPALi analysis. Our results unambiguously demonstrated p44E recombination. The cloning method developed would facilitate precise analysis of the recombination process and the extent of diversity which the recombination creates in the antigenic repertoire.

*Anaplasma phagocytophilum*, previously known as *Ehrlichia* (*Cytopetes*) *phagocytophila*, *Ehrlichia equi*, or human granulocytic ehrlichiosis (HGE) agent, is a tick-borne obligate intracellular bacterium that infects the granulocytes of various mammals, including humans, sheep, goats, horses, dogs, cattle, llamas, and rodents (10). In its mammalian hosts, this rickettsial agent causes acute and chronic febrile illness characterized by hematological abnormalities and immunosuppression. The human manifestation of *A. phagocytophilum* infection is known as human granulocytic anaplasmosis (previously known as HGE) and is potentially fatal (8).

The p44 multigene family of *A. phagocytophilum* encodes immunodominant 44-kDa major outer membrane proteins, P44s (38), that play critical roles in infection. For example, anti-P44 antibodies can protect mice from experimental infection with *A. phagocytophilum* (14), and a recombinant P44 protein induces proinflammatory cytokines in human leukocytes in vitro (15). The p44 gene family has a central hypervariable region of approximately 280 bp. This region is flanked by conserved 5' and 3' terminal sequences that extend in variable lengths of 50 to 500 bp (19, 38). Although each short p44 paralog does not encode an entire P44 protein, by Northern blot analysis we found the full-length p44 transcript containing a hypervariable region identical to that of the short p44 paralog, and by Western blot analysis, we found the full-length P44 protein corresponding to the transcript (38). For this reason, similar to *msp2* of a bovine erythrocytic agent, *Anaplasma marginale* (5), we refer to them as functional pseudogenes to distinguish them from classical pseudogenes that are on the way to elimination.

Using a heterogeneous (noncloned) population of *A. phagocytophilum*, several studies reported that diverse p44 paralogs are expressed in cell cultures, mammals, and ticks (3, 11, 19, 20, 33, 37, 39). Wang et al. demonstrated that rapid switch-off or clearance of the initial dominant transcript p44-18 population occurred during the logarithmic increase of *A. phagocytophilum* in the blood of infected horses. This initial switch-off or clearance event was accompanied by sequential population changeovers of the p44 transcript species concomitant with the development of antibodies to the hypervariable region of the P44-18 protein (33). When *A. phagocytophilum* was preincubated with plasma from an infected horse and then incubated with HL-60 cells, the dominance of the p44-18 transcript was rapidly suppressed in vitro. In addition, most of the newly emerged p44 transcript species were of a type previously undetected in the horse (33). Thus, this antigenic variation system confers the ability to avoid host immune responses and adapt to the new environment.

In a previous study, Barbet et al. identified a polymorphic p44 expression locus consisting of p44ESup1, msp2 (we refer to this gene as p44E to distinguish it from other p44 paralogs and msp2), and the truncated recA in *A. phagocytophilum* strains NY-18, Webster, and HEG2 and in HGE patient number 2 (2).
Lin et al. independently identified a single polymorphic p44 expression locus in strains NY-31, NY-36, NY-37, and HZ of A. phagocytophilum and determined that it consists of four tandem genes: tr1, omp-IX, omp-1N (corresponding to p44ESup1), and p44E (corresponding to msp2) (19). ndk is located upstream of tr1, and truncated recA and valS are located downstream of p44E (19). tr1, omp-IX,omp-1N, and p44E, but not ndk and valS, can be cotranscribed in HL-60 cell culture (19). The sequence alignment of upstream and downstream sequences of p44E in four strains described by Barbet et al. (2) demonstrated that they were 99.6 to 99.9% identical to the respective sequences in four strains identified by Lin et al. (19). These findings suggest that this locus is universally conserved among various A. phagocytophilum human isolates.

\( tr1, \text{omp-IX, and omp-1N of A. phagocytophilum are homologous to tr1, omp-IX, and omp-1N, respectively, of the omp-1 expression locus of Ehrlichia chaffeensis.} \)

We consistently used the p44 designation (38, 40), as a separate ms22 expression locus exists downstream of another two tandem omp-1 homologs within the A. phagocytophilum HZ genome and there are structural differences between P44 and MSP2 (18). A similar polymorphic expression locus, msp2, was identified previously in A. marginale (1). The prevailing hypothesis regarding the mechanism of msp2 and p44 gene expression is that segmental gene conversion of multiple msp2s of A. marginale or p44s of A. phagocytophilum occurs at this single expression locus (1, 6, 7). Gene conversion is defined as a nonreciprocal transfer of genetic information from one molecule to its homolog (30). However, p44 and msp2 recombination could not be proven due to the fact that each A. phagocytophilum or A. marginale strain is heterogeneous with respect to the expression locus sequence. In the present study, therefore, we developed a novel method to clone this obligate intragranulocytic bacterium from the extracellular stage of this bacterial life cycle. Using this cloned population containing a defined p44E sequence, we investigated the temporal changes in p44E composition and sequences in the blood during acute-phase infections in two well-established laboratory animal models: naive equine and severe combined immunodeficiency (SCID) murine hosts.

MATERIALS AND METHODS

Cloning A. phagocytophilum. A. phagocytophilum strain HZ isolated from a human granulocytic anaplasmosis patient in 1995 (32) was cultivated in HL-60 cells at 37°C in 5% CO\textsubscript{2}-95% air using RPMI 1640 medium (Invitrogen, Carlsbad, CA) supplemented with 5% heat-inactivated fetal bovine serum (U.S. Bio-technologies, Parker Ford, PA). The number of passages of heterogeneous strain HZ prior to cloning exceeded 100. The infection rate was 90 to 100% infected HL-60 cells was harvested at 37°C in 5% CO\textsubscript{2}-95% air, using RPMI 1640 medium (Invitrogen, Carlsbad, CA) supplemented with 5% heat-inactivated fetal bovine serum (U.S. Bio-technologies, Parker Ford, PA). A sample of 90 to 100% heat-inactivated fetal bovine serum (U.S. Bio-technologies, Parker Ford, PA). A sample of 90 to 100% infected HL-60 cells was harvested at 37°C in 5% CO\textsubscript{2}-95% air for 6 to 8 h. The peripheral blood leukocytes of each mouse and the horse were prepared immediately after blood collection from these mice, as described elsewhere (15), and stored in RNA later (QIAGEN, Valencia, CA) at \(-20°C\) for later use.

\( p44E \) PCR and sequencing. Total DNA was extracted by using the QIAamp DNA mini kit (QIAGEN). PCR was performed in a 50-μl reaction mixture containing 4 μl of the DNA product, 10 pmol of each primer, 0.2 mM concentration of each deoxynucleoside triphosphate, 5 units of Taq DNA polymerase (Invitrogen), and 1.5 mM MgCl\textsubscript{2}. PCR was performed with 2 min of denaturation at 94°C, followed by 35 cycles of 30 s of denaturation at 94°C, 30 s of annealing at 55°C, and 1 min of extension at 72°C. PCR products were separated by agarose gel electrophoresis and purified from a gel by the QIAEX II gel extraction kit (QIAGEN) and cloned into the pCR-XL-TOPO vector (Invitrogen). Primer pairs used for p44E amplification were IG-1 (5'-GGTTATTTTGCTAGAGAAG-3') and p44ES-3R5 (5'-AACAGATAGCATCGCCAAAAC-3'). The 18-18E* primer pairs were p44-18hsF (5'-GAGACAAGCGGAAAGGTTG-3') and p44-18hsR (5'-AGTCTTATTTGCTAGAGAAG-3'). A total of 20 to 64 DNA clones were randomly selected from the transformants and sequenced on an ABI 373XL Stretch DNA sequencer using the ABI PRISM BigDye terminator cycle sequencing reaction kit (ABI, Foster City, CA).

Sequence and putative recombination breakpoint analysis. To predict the p44 expression site, the sequences of p44E DNA clones from the horse and SCID mouse blood specimens were aligned with the sequences of their respective donor paralogs in the genome, and the aligned sequences were analyzed by the TOPAli software (www.bios.ac.uk/~ianm/topali/) (22–24, 26). The detailed alignment of TOPAli was described elsewhere (22–24). In brief, the pairwise distance matrix of the aligned sequences was calculated using the Jukes-Cantor nucleotide substitution model (10). A window of 50 was moved forward and reverse along a nucleotide sequence alignment at certain stepwise intervals. For each window, the differences of sums of squares of forward (\(D^f\)) and reverse (\(D^r\)) windows were calculated and the corresponding \(D^e\) value was obtained from the maximum (\(D^f\), \(D^r\)). A set of adjacent large \(D^e\) values forming a significant peak suggested a putative recombination breakpoint (22–24). Sequence assembling, alignments, and analysis were performed using the SeqMan, MegaAlign, and MapDraw programs (DNASTar, Inc., Madison, WI).

Nucleotide sequence accession number. The GenBank accession number for p44-18E* (the isogenic clone used in this study) is AJ763488. The GenBank accession numbers for p44s detected in horse EQ006 are AJ763471 to AJ763485, AJ763487, and AJ763489. The GenBank accession numbers for p44s detected in SCID mice are AJ763486 and AJ763490 to AJ763499.

RESULTS

Experimental design. Figure 1 illustrates the experimental design used to study the gene conversion of p44s in the p44 expression locus. In the present study, the A. phagocytophilum
HZ congenic strain was cloned by limiting dilution to obtain an isogenic strain containing a defined p44E sequence. The establishment of the isogenic strain of A. phagocytophilum was confirmed by PCR and/or sequencing of p44E. Next, a naive horse and SCID mice were inoculated with the cloned population. Blood samples were collected at different time points during an approximately 2-month period, and the entire p44E (~1,500 bp) with flanking regions was sequenced in a total of 702 p44E PCR clones.

**Cloning A. phagocytophilum.** Bacteria spontaneously released from HL-60 cells were harvested using a filter and transwell plate, and a single viable rickettsia was identified using the LIVE/DEAD BacLight bacterial viability kit. After three cycles of limiting dilution of spontaneously released host cell-free organisms in HL-60 cells, cloning of the isogenic strain was confirmed by p44E-specific PCR and DNA sequencing of the entire clone in this study was designated p44E-18E.

**p44E conversion in a naive horse and putative recombination breakpoint analysis.** Following inoculation of a naive immunocompetent horse with the cloned A. phagocytophilum strain, the p44E sequences in blood specimens from P.I. days 0 (inoculum), 6, 10, 22, 44, and 58 were determined by p44E-specific PCR and DNA sequencing of the cloned PCR products. A total of 350 randomly selected p44E PCR clones were tested with PCR using p44-18E-specific primers. All 350 clones had the p44E-18E clone in this study was designated p44E-18E.

**TABLE 1.** p44E species and their frequencies at the p44 expression locus in the blood of the horse infected with the cloned A. phagocytophilum

<table>
<thead>
<tr>
<th>p44E species</th>
<th>No. of clones (%) on day:</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>6</td>
</tr>
<tr>
<td>p44-18Eb</td>
<td>40 (90)</td>
</tr>
<tr>
<td>p44-13Eb</td>
<td>2 (4)</td>
</tr>
<tr>
<td>p44-21Eb</td>
<td>2 (4)</td>
</tr>
<tr>
<td>p44-23Eb</td>
<td>1 (2)</td>
</tr>
<tr>
<td>p44-24Eb</td>
<td></td>
</tr>
<tr>
<td>p44-26Eb</td>
<td></td>
</tr>
<tr>
<td>p44-27Eb</td>
<td></td>
</tr>
<tr>
<td>p44-30Eb</td>
<td>6 (13)</td>
</tr>
<tr>
<td>p44-35Eb</td>
<td></td>
</tr>
<tr>
<td>p44-36Eb</td>
<td></td>
</tr>
<tr>
<td>p44-51Eb</td>
<td>2 (5)</td>
</tr>
<tr>
<td>p44-53Eb</td>
<td>8 (20)</td>
</tr>
</tbody>
</table>

* Corresponding donor complete p44: p44-23 (1,188 bp), p44-24 (1,194 bp), p44-36 (1,101 bp), and p44-53 (1,260 bp).
* Number of different p44E species found on each of the following days: 6, 10, 22, 44, 58. 1. Number of total PCR clones analyzed on each of the following days: 6, 10, 22, 44, 58.

**FIG. 1.** Experimental design. (A) Schematic representation of the genetic region of the p44 expression locus. (B) A. phagocytophilum cloning by limiting dilution. Infection of a horse and SCID mice with the cloned A. phagocytophilum that possesses p44-18E. PCR amplification and sequencing analysis of the p44 exchanges in p44E were performed with primers specific to the expression locus. I.V., intravenously; I.P., intraperitoneally.
ulation consisted of six p44E species, p44-24E, p44-26E, p44-27E, p44-30E, p44-36E, and p44-53E, without any single p44E species being dominant (Table 1). On P.I. day 44, p44-18E was the dominant p44E species, representing 78% of the population. However, sequence alignment of the 50 p44-18E clones from P.I. day 44 with the original p44-18E* revealed that there were three types of p44-18E sequences, all of which had several base pair differences in the flanking conserved regions (Fig. 2B), indicating that these were the recombinant p44-18E populations rather than the p44-18E* population contained in the original inoculum. A new p44E species, p44-35E, was detected at P.I. day 44, representing 20% of the total population. At P.I. day 58, only a single p44E species, p44-35E, was detected in the population (Table 1). The noncoding flanking regions of p44E of the 242 p44E clones that were sequenced were identical. These data indicate that 11 new p44E sequences distinct from p44-18E*, constituting 48% (115/242) of p44E sequences and not including three types of p44-18 variants, were detected during the 58-day infection period. Degrees of reliability of percentages of p44E populations were calculated by binomial distribution analysis (13). On day 6, the proportion of p44-23E was 2%, with the sample size of 45. This means that we have a 99% confidence level to conclude that the proportion of p44-23E was between 0% and 7.6%. On day 10, the proportion of p44-53E was 20%, with a sample size of 45, and we have a 99% confidence level to conclude that the proportion of p44-53E was between 3.9% and 36.1%. The proportion of p44-18E was 82% on day 10, with the sample size of 45, and we have a 99% confidence level to conclude that the proportion of p44-18E was between 67.4% and 97.0%. Thus, the percentage is relatively less accurate at lower percentages of p44E populations. However, even considering confidence limits, lack of detection of any p44-18E population at P.I. days 22 and 58 was a significant reduction.

For every p44E sequence detected in the horse blood specimens, putative donor p44s with the identical hypervariable region sequence were found elsewhere in the A. phagocytophilum HZ genome, both in our previous studies (3, 11, 19, 20, 33, 37, 39) and by BLAST search of the A. phagocytophilum genome sequence database (www.tigr.org). As shown in Table 1, 77% (19/24) of the predicted donor p44s for different p44E species detected in the horse were functional pseudogenes based on the truncation in the 5’ and/or 3’ end and the absence of common translational start codons (17).

TOPALi is graphical analysis software that uses statistical analysis to detect the presence of possible recombination sequences and gives an approximate location of the breakpoints within DNA multiple alignments. The program checks a multiple alignment for evidence of recombinant sequences (24). The performance of the method was previously assessed in a simulation study and proven successful in detecting a variety of recombination events (24). The sequences of the recombinant p44-13E, p44-21E, and p44-23E species isolated at P.I. day 6 of each of the putative donor p44s and of the recipient p44-18E* were aligned, and the putative recombination points were analyzed by TOPALi (Fig. 2A). Using TOPALi between p44-18E* and p44-13, the recombination events were predicted at 301, 382, 1,023, and 1,130 bp in p44-13E (Fig. 2A[i]). The sequences between the putative breakpoints, including the hypervariable region, were identical between donor p44-13 and p44-13E (Fig. 2A[ii]). The p44-21 gene conversion event might have occurred approximately at 401, 497, 860, and 949 bp in the flanking conserved regions of p44-21E (Fig. 2A[iii]). For the p44-23 gene conversion event, two sharp putative recombination breakpoints were detected at approximately 340 and 490 bp in the 5’ end flanking the conserved region, and a broad putative recombination region was detected around 935 bp in the 3’ end flanking the conserved region of p44-23E (Fig. 2A[iii]). On P.I. day 44, three types of p44-18E were detected that were distinct from p44-18E* at the border of the hypervariable and conserved regions (Fig. 2B). To determine whether sequential recombination of multiple p44s at the recipient p44 expression locus creates the mosaic sequence at the border regions as previously proposed (19), the sequences of all p44E species on P.I. days 6, 10, 22, and 44 (representing possible recipient p44 sequences) were aligned with both the donor p44-18 and the three types of p44-18Es that were detected at P.I. day 44 and analyzed by TOPALi. Our results revealed that the p44-18Es on P.I. day 44 consisted of five segments, which could have been derived from p44-18E*, p44-51, p44-24E, p44-26E, and the donor p44-18 (Fig. 2B). By determining p44E sequences that had changed in the cloned A. phagocytophilum population, we, for the first time, unambiguously demonstrated p44 gene conversion in the expression locus. This gene conversion event generated point mutation-like sequence variations and mosaic sequences at a putative recombination crossover region within 5’ and 3’ conserved regions of p44E.

p44 changes in SCID mice and putative recombination breakpoint analysis. To investigate the dependence of p44 recombination on acquired host immune responses, SCID mice were inoculated with the cloned A. phagocytophilum strain. Following inoculation, the p44E sequences in blood specimens from P.I. days 0 (prior to inoculation), 5, 10, 15, 20, and 50 were determined by p44E-specific PCR and DNA sequencing of the PCR products (Table 2). The P.I. day 0 inoculum was identical to the batch used to confirm the clone as described above, and 100% of the population consisted of p44-18E. On P.I. day 5, 100% of the p44E detected was p44-18E* in both SCID mouse no. 1 and 2. The p44E population varied in individual mice after P.I. day 10. On P.I. day 10, p44-18E* was replaced completely by p44-51E in mouse no. 3, while in mouse no. 4, p44-18E* remained the major p44 species, but six new p44Es, p44-24E, p44-30E, p44-48E, p44-51E, p44-26E, and p44-13E, had recombined into the p44 expression locus. On P.I. day 15, a p44-18E species was the major p44 species (72% in mouse no. 5 and 55% in mouse no. 6). On P.I. day 20, p44-18E remained above 50% in mouse no. 7 (100%) and mouse no. 8 (50%). On P.I. day 50, p44-18E was the major species in mouse no. 9, comprising 87% of p44E clones, whereas in mouse no. 10, p44-18E could not be detected in the p44 expression locus. These data indicate that an acquired immune response is not required for p44E genetic variation to occur. However, p44-18E was more than 50% of the p44E species detected in 8 of 10 SCID mice during the 50-day infection period. During the 50-day infection period, the total of 13 new p44E sequences divergent from the original p44-18E* changed in the blood of SCID mice, constituting 42% (192/460) of p44E sequences. As shown in Table 2, 86% (12/14) of predicted donor p44s for different p44E species in the SCID.
FIG. 2. Recombination breakpoint analysis of aligned sequences of \textit{p44E}s and putative donors by TOPALi. (A) Nucleotide sequence alignment of \textit{p44-13E} [i], \textit{p44-21E} [ii], and \textit{p44-23E} [iii] and putative donors \textit{p44-13}, \textit{p44-21}, and \textit{p44-23}, respectively, at P.I. day 6 from a horse inoculated with the cloned \textit{A. phagocytophilum}, \textit{p44-18E*}. (B) Nucleotide sequence alignment of three \textit{p44-18E}s (T1, T2, and T3) at P.I. day 44 and the putative recipient \textit{p44E}s, \textit{p44-18E*}, \textit{p44-51E}, \textit{p44-24E}, and \textit{p44-26E}, which were detected at previous time points, and the putative donor \textit{p44-18}. Boxed areas are identical sequences among \textit{p44E}s and putative recipient \textit{p44E}s. (C) Nucleotide sequence alignment of \textit{p44-51E} and the putative donor \textit{p44-51} at P.I. day 10 from SCID mouse no. 3 inoculated with the cloned \textit{A. phagocytophilum}. Black boxes represent the putative recombination sites (aligned sequences of these sites are shown under each figure). Hypervariable regions are shown in different colors. The y axis represents the difference of sums of squares (\textit{Dss}). The x axis represents sequence position. Numbers in black indicate the putative recombination breakpoint positions. The red horizontal dotted line is the 95% significance point of \textit{Dss}.
FIG. 2—Continued.
mice were functional pseudogenes based on the truncation in 
5′ and/or 3′ end and the absence of the common transla-
tional start codon (17), which is greater than the percentage of func-
tional pseudogenes identified in the horse (67%, 8/12).

TOPALI analysis of the new p44E sequences that were de-
tected on P.I. day 10 in SCID mouse no. 3 revealed recombi-
nation breakpoints similar to those found in the analysis of horse 
species (Fig. 2C). Putative recombination break-
points were detected approximately at 371 and 883 bp for 
p44-51E conversion.

DISCUSSION

The present study describes the successful cloning of A. 
phagocytophilum. Conventional bacterial cloning techniques 
cannot be used for A. phagocytophilum because it does not 
 grow on agar plates. While the plaque cloning method was 
 successfully used for cloning Rickettsia species in cell cul-
ture (21, 28, 31, 34), A. phagocytophilum does not make distinct 
plaques. Moreover, this bacterium grows as tight clusters of 
colonies in membrane-bound inclusions called “morulae,” 
and bacteria within a single host cell may consist of 
multiple clones with differing p44E. The only stage during the 
A. phagocytophilum life cycle in which these rickettsiae are 
dispersed as individual particles appears to be during sponta-
neous release from host cells (H. Niu and Y. Rikihisa, unpub-
lished data). Furthermore, mechanical or physical stress, such 
as sonication and homogenization, to release intracellular bac-
teria appears to enhance recombination (X. Wang and Y. 
Rikihisa, unpublished data). These biological characteristics of 
A. phagocytophilum should be taken into consideration for 
future cloning and maintenance of cloned organisms.

By using this cloned population, we demonstrated the re-
combination in the p44 expression locus of A. phagocytophilum 
in two animal infection models. Although it is impossible to 
determine the exact time when the recombination took place, 
11 new p44E sequences were detected in the horse and 13 new 
p44E sequences were detected in a total of 10 SCID mice.

Overall, 48% and 42% of the p44E population in the horse and 
in SCID mice, respectively, were p44E species distinct from 
p44-18E. Since A. phagocytophilum becomes undetectable in 
immunocompetent mice by approximately 2 to 3 weeks, it is 
currently impossible to determine the long-term p44E changes 
in these mice. However, comparison of the horse and SCID 
mice data showed that levels of p44E converts were similar in 
either the absence or presence of selection by the acquired 
immune system. This is consistent with our previous work that 
demonstrated within-host p44E and p44 transcript species 
experimentation in horses occurring before antibodies selec-
ting against the activated P44s had developed (33). These 
data suggest that the host innate immune system has a signif-
icant role in p44E recombination and/or selection. Selection or 
adaptation of the p44E population also occurs in cell culture 
(20). It is currently impossible to compare p44E recombination 
rates between animals and in cell culture, since once recombi-
nation occurs, we could not prevent selection or adaptation.

Hence, the greater levels of p44E variants in animals than in 
cell culture, as seen previously (33), may simply be due to 
higher levels of selection in animals than in cell culture. How-
ever, we could not deny the possibility that unknown in vivo 
selective factors enhanced the p44E recombination rate in animals. It 
is well known that vlsE recombination in Borellia burgdorferi 
does not occur in culture but is induced in mammals by in vivo 
factors (36).

Our previous cell culture study showed that immune plasma 
modulates p44 population dynamics (33). In the infected horse, 
the genetic species p44-18E* remained on P.I. day 6 (90% of 
the total population), suggesting that this population repli-
cated without obvious recombination at the p44 expression 
locus at least until P.I. day 6. This genetic species became 
undetectable at P.I. day 22. Although p44-18E became the 
dominant species at P.I. day 44, the fact that the sequence of 
this p44-18E was different from the p44-18E* indicates that this 
was not selective growth of the original p44-18E* species. Since
infected horses develop antibodies specific to the P44-18 hypervariable region (33, 39), the fact that a dominant population in the horse reverted to p44-18E suggests that the B-cell immunity specific to the P44-18 hypervariable region could not last long enough to prevent the emergence of p44-18Es. Alternatively, it is possible that base changes at the borders of the hypervariable region of p44-18E sufficiently affect protein folding or important epitopes to generate escape variants.

Intriguingly, next to p44-18E, the most frequently observed p44Es in the horse and mice were p44-31E, p44-35E, and/or p44-30E. A BLAST search of the A. phagocytophilum genome sequence demonstrated that the donors p44-31, p44-35, and p44-30 were arranged in tandem 293 bp upstream of the full-length recA. This 293-bp sequence is almost identical to the sequence between p44E and its downstream truncated recA in the p44 expression locus (except for a 1 bp deletion in the intergenic space between p44-30 and recA 102 bp downstream of p44-30). This suggests that this 557-bp region including the 293-bp intergenic region and the entire 264 bp of the truncated recA in the expression locus is duplicated in the p44-31-35-30 locus and perhaps that this repeat sequence is involved in RecF-dependent recombination and absence of homologous recombination. To recombine, various reserve short and full-length p44 and p44 pseudogenes recombined to the expression locus and that the majority of recombinant p44Es were p44 functional pseudogenes. The possible factors that affect this apparent recombination preference are currently unknown.

While in this study a maximum of seven different p44 hypervariable species were detected in the p44 expression locus in the blood of infected mammals at a given time point, more diverse expression is expected in nature due to the combined result of multiple clones recombining with their preferred p44Es, as suggested by our previous reports on the p44 cDNA population of tick-transmitted noncloned A. phagocytophilum in mice and in human patients (11, 20, 39).

TOPALI analysis in the present study predicted variations in recombinant regions and region length (from 50 to 200 bp) in the 5' and 3' conserved regions, which may account for the sequence variations in p44 cDNA concentrated within 50 to 200 bp at the border of the hypervariable region (11, 19). This analysis is consistent with the previous prediction that instead of using the entire conserved region for gene conversion, various lengths (50 to 200 bp) of partial 5' and 3' conserved region sequences at the border of the hypervariable region are used for gene conversion in the p44 expression locus (19). Furthermore, no mixing sequences external to the putative recombination region were observed in our 702 sequenced p44Es, suggesting that heteroduplex formation is limited to the border regions. The noncoding flanking regions of p44E species were identical, consistent with the concept of gene conversion occurring without crossing over. In homologous recombination, base pairing leads to the alignment of relatively long homologous sequences (>50 bp), and homologous recombination involves degradation and synthesis of DNA (16). However, in site-specific recombination, short pieces of homologous DNA (a few 10’s of base pairs) present on both the incoming and the resident DNA molecule, are cut in both strands at two specific positions (9, 12). The length and structure of the breakpoint regions which were detected in the present study were compatible with the concept of homologous recombination rather than site-specific recombination.

In the present study, p44 donors are composed of both full-length p44s and short pseudogenes that function as cassettes to provide the entire p44E hypervariable region and are thus distinctly different from the gene conversion of multiple segments within the hypervariable region, as described for Neisseria gonorrhoeae pilS/pilE (25), Borrelia burgdorferi vlsE (35), Anaplasma marginale msp2 (1, 2, 4, 7), and Mycoplasma synoviae v1hA (27). Although we cannot deny the possibility of p44E segmental gene conversion under conditions different from our study, previously described segmental gene conversion in msp2 (p44) of A. phagocytophilum (2) might be explained by variable mosaic sequences within 3' and 5' conserved regions, as detected in the recombination breakpoint analysis.

Previous studies showed that p44s were maintained in their donor loci (19, 38), suggesting gene conversion without crossover of flanking loci. p44 pseudogene loci, in particular, are maintained in a silent state, preventing their transcription; however, they can serve as donors of genetic information at p44 by recombination. Using a BLAST search of the 1.5-Mb A. phagocytophilum genome, we found that all 16 newly converted p44Es in the blood of the infected horse and SCID mice in the present study had a hypervariable region sequence identical to that of the putative p44 donors between 78 kb upstream and 120 kb downstream of the p44E locus within 200 kb, except p44-21 and p44-34. These data suggest that the presence of p44s near the p44E locus would facilitate the homologous recombination. To recombine, various reserve short and full-length p44s in the same genome need to physically interact with the recipient p44E. Based on the presence of genes involved in RecF-dependent recombination and absence of genes involved in RecBCD-dependent recombination in the A. phagocytophilum HZ genome, we previously proposed that RecF-dependent recombination requiring replication was the mechanism of p44 recombination (19). Initiation of chromosomal replication would provide partially duplicated sister chromosomes within a single bacterial cell. We speculate, thus, that p44 recombination takes place between sister chromosomes rather than intrachromosomally. This allows the loss of the donor chromosome while ensuring that the recipient chromosome would survive, keeping the same chromosome structure. Therefore, p44E homologous recombination and, consequently, antigenic variation are the costly endeavors in which A. phagocytophilum readily sacrifices its own growth to survive in the mammalian host.

ACKNOWLEDGMENTS

We thank X. Wang and T.-H. Lai for assistance with the horse experiment. The genome of A. phagocytophilum was sequenced at The Institute for Genomic Research, Rockville, MD. Sequences are available at http://www.tigr.org.

This research was supported by National Institutes of Health grant R01 AI47407 to Y.R. The sequencing project was supported by National Institutes of Health grant R01 AI47885 to Y.R.

REFERENCES


