Groupings of highly similar major surface protein (p44)-encoding paralogues: a potential index of genetic diversity amongst isolates of *Anaplasma phagocytophilum*

A. N. J. Casey,1 R. J. Birtles,2 A. D. Radford,3 K. J. Bown,1 N. P. French,3 Z. Woldehiwet2 and N. H. Ogden1

Correspondence
N. H. Ogden
ogdenn@courrier.umontreal.ca

1Department of Veterinary Preclinical Science, University of Liverpool, Brownlow Hill and Crown Street, Liverpool L69 7ZJ, UK
2,3Department of Veterinary Pathology and Department of Veterinary Clinical Science, University of Liverpool, Leahurst, Neston, South Wirral CH64 7TE, UK

*Anaplasma phagocytophilum* is a tick-borne bacterium that is zoonotic in the USA and southern Europe, but although the bacterium is endemic in the UK, no cases of clinical human disease have yet been detected in that country. Potential genomic differences amongst UK and USA isolates were investigated by comparing partial 16S rRNA gene and p44 paralogue sequences amplified by PCR from 10 UK ruminant or tick isolates, with published sequences from USA isolates. No significant clustering among the isolates was resolved by phylogenetic analysis of alignments containing 16S rRNA gene sequences. The structure of predicted proteins encoded by p44 paralogues, amplified from 81 clones obtained from the UK isolates, was similar to that described previously for paralogues from USA isolates. Paralogue sequences did not obviously cluster by country, host species or isolate, but most paralogues were 30–70% similar, making meaningful alignments difficult. Some p44 paralogues from different isolates formed clusters of sequences that were more than 90% similar to one another (‘similarity groups’). The paralogues in each cluster were particularly similar in gene regions most likely to code for ligands. In the sample studied, 95% of the similarity groups comprised paralogues from either USA or UK isolates only and occurred with greater frequency amongst paralogues from USA rather than UK isolates. These findings raise the hypothesis that sequences of paralogues in similarity groups may provide an index of adaptation of different ‘strains’ of *A. phagocytophilum* to specific reservoir hosts in different geographical locations, and any associations with infectivity for different species including humans.

**INTRODUCTION**

*Anaplasma phagocytophilum* is a tick-borne bacterium that has been recognized as a pathogen of domesticated animals for many years (Woldehiwet & Scott, 1993). It has recently been recognized that this bacterium is a zoonosis causing potentially fatal illnesses, which occurs in humans in the USA (Chen et al., 1994) as well as in southern Europe (Lotric-Furlan et al., 1998). It has been known for many years that *A. phagocytophilum* is widespread and common in *Ixodes ricinus* ticks in UK habitats, particularly in habitats where the ticks and bacteria are maintained by sheep (Woldehiwet & Scott, 1993). To date, however, there are not known to have been any cases of disease in humans associated with *A. phagocytophilum* in the UK even though there is some serological evidence for the occurrence of human infections there (Sumption et al., 1995; Thomas et al., 1998). This could suggest that *A. phagocytophilum* populations maintained in the UK, southern Europe and particularly the USA have different characteristics in terms of their pathogenicity in humans, implying that clinically significant genetic heterogeneity may occur amongst isolates of this bacterium.

*A. phagocytophilum* [formerly named as the agent of human granulocytic ehrlichiosis (HGE), *Ehrlichia phagocytophila* and *Ehrlichia equi*; Ristic & Huxsoll, 1984; Dumler et al., 2001] can infect a very wide range of species (wild rodents...
and deer, cats, dogs, llamas and humans: Ogden et al., 1998) and some isolates can cross-infect between species (Foley et al., 2002). Some isolates from one species do not seem to be directly infective for another (Pusterla et al., 2001), however, and early cross-protection studies using European isolates from sheep (i.e. previously E. phagocytophila), suggested that a high degree of 'strain' diversity may occur amongst different isolates of A. phagocytophilum (Foggie, 1951; Tuomi, 1967).

Such conflicting observations could in part be explained by the characteristics of the immunodominant 44 kDa protein (P44) of A. phagocytophilum, which is encoded by a multigene family (p44; Murphy et al., 1998). Expression of different paralogues occurs in different environments such as mammalian host and vector tick cells (Zhi et al., 2002a, b), and differential expression is also thought to be responsible for antigenic variation in the host (Barbet et al., 2003). The discovery of this property of the bacterium raises the hypothesis that evidence for 'strain' diversity amongst A. phagocytophilum isolates observed in some experiments could have been due to phenotypic rather than genetic variations. Nevertheless, the close interaction of expressed p44 proteins with host cells (Park et al., 2003) may mean that some co-evolution of this gene with different reservoir hosts in different foci (e.g. rodents in the USA as opposed to sheep in the UK) has occurred. This could have consequences for the potential infectivity of A. phagocytophilum, in infected ticks in different foci of infection, for humans or domesticated animals.

Some studies have already suggested that variations in either 16S rRNA gene or p44 sequences from USA isolates may be associated with ecological or clinical characteristics of A. phagocytophilum (Carter et al., 2001; Massung et al., 2002). In the present study, we have investigated the hypothesis that differences in either 16S rRNA gene or p44 sequences may occur amongst UK isolates from ruminants (that have yet to be associated with human disease), and isolates from other parts of the world, particularly the USA, where human disease does occur.

**METHODS**

**UK 'strains' and other isolates used in the study.** Six ovine 'strains' and one caprine 'strain' of A. phagocytophilum were used in the study; these had been previously characterized in cross-protection studies, and maintained in the laboratory as blood stablulates, as previously described (Woldehiwet, 1981). Briefly, stablulates of heparinized infected whole sheep blood obtained on the second day of bacteraemia were stored at −114 °C, with 10% dimethyl sulphoxide as cryopreservative. Before use, cryopreserved stablulates were rapidly thawed and diluted 1 in 10 with sterile phosphate-buffered saline (PBS), pH 7-2 (Woldehiwet, 1981). One millilitre of diluted stablulate of each isolate was inoculated intravenously into one susceptible sheep bred and maintained in a tick-free environment. Blood samples were used to monitor infection by microscopy of Giemsa-stained smears and PCR (Ogden et al., 2002). DNA extracted from blood samples collected during the second day of bacteraemia was used for the amplification and sequencing of the 16S rRNA gene and the p44 gene as described below. The 'strains' used were those originally isolated from sheep in different regions of Scotland and named 'Aberfeldy' (AB), 'Cairn', 'Harris', 'Lephimore' and 'Old Soughope' (OS), and the 'Feral Goat 'strain' originally isolated from a goat in Scotland. We also used DNA extracted from the blood of a naturally infected sheep in North Wales, an engorged nymphal Ixodes ricinus tick collected from a different sheep in North Wales, and from blood collected from a cow in Somerset (SW England) that had shown signs of acute 'pasture fever' (Woldehiwet & Scott, 1993) and in which A. phagocytophilum infection had been diagnosed by microscopy of blood smears.

**Amplification and analysis of 16S rRNA gene sequences.** DNA was extracted from 200 µl of each blood sample used in the study using DNeasy blood and tissue kits (Qiagen). DNA was extracted from the engorged tick by alkaline digestion as previously described (Ogden et al., 2002). A nested PCR was used to amplify a 546 bp region of the 16S rRNA gene of A. phagocytophilum (Massung et al., 1998). Negative controls were incorporated at a rate of one per test sample. Purified PCR products (Wizard PCR Preps DNA Purification system; Promega) were directly sequenced with primers ge9f and ge2 (Massung et al., 1998) according to the manufacturer's instructions (ABI prism dye terminator cycle sequencing ready reaction kit; Perkin Elmer). 16S rRNA gene sequences from this study were compared to published sequences of 19 isolates of A. phagocytophilum from Europe, USA and Asia, and to sequences from related bacteria. All sequences were aligned using the PHYLIP program of GCG (Wisconsin Sequence Analysis Package, Genetics Computer Group, version 8.1). Phylogenetic analysis of the aligned sequences was performed using the programs NEIGHBOR, DNADIST (Kimura) and SEQBOOT from the PHYLIP package (Felsenstein, 1989). Trees were drawn using TREEVIEW (Windows version 32; Page, 1996).

**Amplification and analysis of p44 sequences.** Amplicons of p44 gene paralogues of approximately 550 bp were obtained from the DNA samples using primers P3708 and P4257, which anneal to the N- and C-terminal conserved regions of p44 paralogues in USA isolates, flanking the central hypervariable region of the genes, as previously described (Zhi et al., 1999). PCR products were cloned into the pcRII plasmid and transformed into Escherichia coli using a pCR 2.1-TOPO TA cloning kit (Invitrogen). Plasmids containing appropriately sized inserts (400–600 bp) were identified using standard protocols, purified (Wizard Plus SV Miniprep DNA Purification System; Promega) and their inserts were sequenced using primers T4 and SP7. The DNA sequences obtained were aligned in GCG with all (65 in number) of the available, published p44 parologue gene sequences of A. phagocytophilum (Caspersen et al., 2002; Lin et al., 2002; Zhi et al., 1999, 2002b), including those of the Webster and BDS 'strains', and the NY31, NT36, NY37, HZ and LL isolates from the USA (GenBank accession nos AF443396 to AF443413 and AF443415 to AF443419, AF059181, AF135263, AF135254 to AF135257, AF412818 to AF412831, AF414591 and AY064513 to AY064530).

A number of comparisons of 'relatedness' amongst p44 paralogues from the UK and USA isolates were made. First, variations in similarity of different parts of the sequenced UK paralogues were investigated using the PHYLIP and PLOTSIMILARITY programs in GCG, and a consensus sequence of paralogues from UK isolates was generated. Antigenicity and hydrophobicity of different parts of this consensus sequence were investigated using the PLOTSTRUCTURE program in GCG for a qualitative comparison with p44 paralogues from USA isolates (Lin et al., 2002). Second, the percentage similarity of each UK sequence obtained in the present study, to each other and to published p44 sequences from USA isolates, was estimated from the DISTANCES program in GCG. From these, uncorrected distance trees of DNA sequences were developed. Due to a very high number of insertions...
and deletions, alignments of these sequences for phylogenetic analyses that could yield statistical support for tree topology were impossible (see Results). We did, however, investigate whether any paralogues had particularly high similarity (> 90%) to any others and whether such similar paralogues were from isolates from the same or different countries. Third, we investigated the sites of base substitutions amongst paralogues that were more than 90% similar to another, using the web-based SNAP software (http://www.hiv.lanl.gov; Korber, 2000). The numbers of non-synonymous substitutions per base in the N- and C-terminal conserved regions of these highly similar paralogue pairs or groups were estimated and then compared with the numbers of non-synonymous substitutions per base in the central hypervariable regions using the non-parametric Mann–Whitney U test.

GenBank accession numbers

Partial 16S rRNA sequences. North Wales sheep isolate AY149635, North Wales tick isolate AY149637, AB 'strain' AY176586, Cairn 'strain' AY176587, Feral Goat 'strain' AY176588, Harris 'strain' AY176589, Lephimore 'strain' AY176590 and OS 'strain' AY176591.

Partial p44 sequences. AB 'strain' paralogue ABI AY176512, Cairn 'strain' paralogues Cairn1, Cairn2, Cairn3, Cairn4, Cairn6, Cairn7, Cairn8, Cairn9, Cairn10, Cairn11, Cairn12, Cairn13, Cairn15, Cairn16 and CairnB01 AY176513 to AY176528 respectively. Bovine isolate paralogues Cow1, Cow2 and Cow3 AY176529 to AY176531 respectively. Feral Goat 'strain' paralogues FGA1, FGA3, FGA4, FGA5, FGA6, FGA7, FGA8, FGA12, FGA13, FGA14, FGA15, FGA17, FGA18, FGA19, FGA20, FGA21, FGA22, FGA23, FGA24, FGA25, FGA26, FGA27, FGA28, FGA29, FGA30, FGA31, FGA32, FGB1, FGB2, FGB4, FGB6, FGB10 and FGB11 AY176532 to AY176564 respectively. Lephimore 'strain' paralogues Lep1, Leph2 and Leph4 AY176565 to AY176567. North Wales sheep isolate paralogues NW2, NW3, NW5, NW7, NW8, NW9, NW11, NW12, NW13, NW14, NW16, NWB1 and NWB2 AY176568 to AY176586 respectively. North Wales tick isolate paralogue NWI, AY176582.

RESULTS

16S rRNA gene sequences

All the sequences obtained in this study were between 98% and 100% similar with published sequences. The sequences obtained from a field-infected sheep in North Wales and from the laboratory 'strains' OS, Lephimore, Cairn and AB were identical to one another, and to sequences from isolates from a human and a llama from the Western USA, and an Ixodes persulcatus tick from Korea (Fig. 1). The sequences obtained from a field-infected tick in North Wales, and the laboratory 'strain' Harris were identical to one another but differed from the sequence of 'strain' OS in having a guanine rather than an adenine at positions 77 and 84 (using the 16S rRNA numbering of Chen et al., 1994). The sequences from the Harris 'strain' and the North Wales tick were identical to those of an isolate from a red deer in Slovenia (GenBank accession no. AF481855; Fig. 1). The sequence obtained from the laboratory 'strain' Feral Goat was identical to the OS 'strain' except for an adenine rather than a guanine at position 468. Apart from one sequence obtained from an I. persulcatus tick in China (GenBank accession no. AF205140), phylogenetic analysis did not reveal any significant clustering of sequences associated with the geographical location or the species of origin of the isolates (Fig. 1).

Fig. 1. Phylogenetic tree based on Kimura distances of 16S rRNA gene sequences of isolates of the A. phagocytophilum genogroup with other members of the α-Proteobacteriaceae by neighbour-joining distance matrix. Sequences obtained in this study are underlined. Numbers at nodes are bootstrap values in support of tree topology based on 100 resamplings. Evolutionary distances are to scale.
p44 gene sequences
A total of 81 p44 clones were sequenced, yielding 68 sequences that differed by more than two bases. Sequence lengths varied from 410 to 503 bp excluding primers. From the Feral Goat and Cairn 'strains', and North Wales field isolate, 37, 16 and 18 clones, respectively, were sequenced, of which 31, 16 and 11, respectively, differed by more than two bases. A further 10 clones were obtained from the isolate from the cow, and the ovine 'strains' Lephimore and AB. There were no identical sequences

**Table 1.** Groups of p44 paralogue DNA sequences with similarity of greater than 90%

The members of each group are indicated in rows 1–27. Numbers in parentheses refer to the identification numbers of paralogue sequences used in GenBank.

<table>
<thead>
<tr>
<th>Location:</th>
<th>USA</th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th>UK</th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Isolate:</td>
<td>HZ</td>
<td>Webster</td>
<td>BDS</td>
<td>NY31</td>
<td>NY36</td>
<td>NY37</td>
<td>Cairn</td>
<td>Feral Goat</td>
<td>Lephimore</td>
<td>North Wales</td>
</tr>
<tr>
<td>Host:</td>
<td>Human</td>
<td>Human</td>
<td>Horse</td>
<td>Human</td>
<td>Human</td>
<td>Human</td>
<td>Sheep</td>
<td>Sheep</td>
<td>Sheep</td>
<td>Sheep</td>
</tr>
<tr>
<td><strong>Similarity group</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>+ (1)</td>
<td>+ (1)</td>
<td>+ (1)</td>
<td>+ (1)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>+ (2)</td>
<td>+ (2)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>+ (3)</td>
<td>+ (13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>4</td>
<td>+ (5)</td>
<td>+ (15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>5</td>
<td>+ (6)</td>
<td>+ (15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>+ (7)</td>
<td>+ (16)</td>
<td>+ (1)</td>
<td>+ (7)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>+ (8)</td>
<td>+ (17)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>8</td>
<td>+ (10)</td>
<td>+ (4)</td>
<td>+ (10)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>9</td>
<td>+ (12)</td>
<td>+ (57)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>+ (13)</td>
<td></td>
<td>+ (13)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>+ (14)</td>
<td>+ (14)</td>
<td>+ (14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>+ (15)</td>
<td>+ (3)</td>
<td>+ (15)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td></td>
<td>+ (23)</td>
<td>+ (23)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>+ (24)</td>
<td>+ (24)</td>
<td>+ (24)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>15</td>
<td>+ (6)</td>
<td>+ (8)</td>
<td>+ (28)</td>
<td>+ (28)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>16</td>
<td>+ (45)</td>
<td></td>
<td>+ (29)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>+ (1)</td>
<td></td>
<td>+ (39)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>18</td>
<td>+ (5)</td>
<td>+ (9)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>19</td>
<td>+ (10)</td>
<td>+ (14)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>20</td>
<td>+ (18)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>21</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>22</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>23</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>24</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>25</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>26</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>27</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
obtained from different 'strains' or isolates and sequences of all but four of the clones were in-frame.

All in-frame sequences, as well as the consensus, yielded putative protein sequences similar to those of sequences from USA isolates, i.e. two moderately hydrophobic N- and C-terminal conserved regions of relatively low potential antigenicity flanking a more hypervariable, potentially hydrophilic and antigenic region that comprised three domains separated by highly conserved amino acids. The highly conserved amino acids that delineated the hypervariable region domains were the same as those identified in USA isolates (Lin et al., 2002). In general, N- and C-terminal conserved regions were those most highly conserved amongst paralogues from all UK and USA isolates. These results are summarized in Fig. 2.

All UK p44 sequences from this study, and published sequences from the USA, were more than 30% similar, but most (99.5% of pairwise comparisons) were less than 79% similar, and no paralogue sequences from UK isolates were 100% similar with any sequences from USA isolates or other UK isolates. A small proportion of sequences (0-3% of pairwise comparisons) formed groups with greater than

Fig. 3. Unrooted Kimura distance tree of DNA sequences of p44 paralogues, from this and other studies. 'Similarity groups' of paralogues with greater than 90% similarity are encircled. A similarity group with members from isolates from the UK and the USA is arrowed. The scale bar represents a corrected 10% peptide difference.
90% similarity (Table 1, Fig. 3) and these 27 groups are termed ‘similarity groups’ in the following. In only two cases were nearest neighbouring sequences more than 79% similar to a member of a similarity group. Six of the ‘similarity groups’ comprised paralogues from the same isolate, but these were only found in the UK isolates Feral Goat and Cairn, none being seen in USA isolates. Some comprised paralogues from different isolates from the same country (two from UK and 19 from USA isolates). Twenty-six of the 27 similarity groups (96%) comprised paralogues from the same country only, and of the 60 paralogues that belonged to a similarity group, 57 (95%) belonged to a group comprising members from the same country only. Much of this difference was due to a difference in the likelihood that different isolates from the same country contained highly similar paralogues: 35/65 (54%) paralogues from USA isolates had >90% similarity with a paralogue in a different USA isolate, while 4/68 (6%) paralogues from UK isolates had >90% similarity with a paralogue in a different UK isolate. One group containing highly similar paralogues from UK and USA isolates comprised the p44-18 paralogue from the HZ ‘strain’ isolated from a human in the USA (Zhi et al., 2002b), and the paralogues FGB6 and Cairn7 from the UK Feral Goat and Cairn ‘strains’, respectively (Fig. 4).

In comparisons of sequences within each ‘similarity group’ there were no inserted or deleted sequences in the central domains amongst members of the same similarity group, variations here being due only to base substitutions (e.g. Fig. 4). Insertions were found in only one sequence in each of three similarity groups, which in each case was a single inserted amino acid in the C-terminus conserved region. For comparisons of the proportions of synonymous base substitutions along the length of the gene in these groups, the inserted amino acid was edited out. There was a significantly greater number of non-synonymous substitutions per base in the N- and C-termini (mean 0.127, SE 0.047) than in the central region (comprising Domains 1, 2 and 3 as defined by Lin et al., 2002) bounded by the absolutely conserved N-terminus cysteine and the C-terminus alanine residues (mean 0.023, SE 0.010, Mann–Whitney \(U=45, P=0.024\); Fig. 5). Therefore, the pattern of variation in conserved and hypervariable regions along the p44 gene, amongst paralogues of the same similarity group, was the inverse of that seen when all of the p44 sequences were compared together (Fig. 2).

**DISCUSSION**

The partial 16S rRNA gene sequences of the UK ‘strains’ and isolates used in this study were highly similar to sequences obtained from other studies in Europe, the USA and Asia. Phylogenetic analysis of this region of the gene failed to resolve any well-supported divergence among the ‘strains’ examined, including those of UK sheep isolates and isolates associated with disease in humans, even though variations in this region of the gene have been reported to have ecological significance (Massung et al., 2002).

Sequences of p44 genes, which were similar to those from USA isolates, were amplified from all of the tested UK ‘strains’ and isolates. Up to 20 paralogues of the p44 gene have been discovered in the USA isolates that have been examined in detail to date (Caspersen et al., 2002). We did
not exhaustively sequence the UK isolates but in one isolate, only six sequenced clones were identical amongst 37, indicating that it most likely had a higher number of p44 paralogues. Some UK *A. phagocytophilum* isolates could, therefore, have a larger complement of different paralogues than their USA counterparts. This phenomenon may have been due to the occurrence of mixed infections in sheep with different 'genotypes' of *A. phagocytophilum* as detected in other studies by 16S rRNA gene sequence analysis (Stuen et al., 2002). Mixed infections could have occurred, however, in the USA isolates studied to date. Similarities amongst and between paralogues from USA and UK isolates were mostly maintained at the putative peptide level and only a very low number of paralogues were out of frame, suggesting that functional integrity of a large number of possible p44 proteins is important for bacterial survival.

Most p44 sequences had only moderate similarity to one another (75% or less) and there was no clear clustering of variants from individual UK and USA isolates, or between isolates, in uncorrected distance trees. Some sequences did, however, form groups with greater than 90% similarity (the ‘similarity groups’), comprising in some cases paralogues from the same isolate, in some cases different isolates from the same countries and host species, and in one case, isolates from different countries and host species. Much of the variation amongst paralogue sequences was due to insertions and deletions of variable size (mostly in the hypervariable domains), which made impossible the meaningful alignments required for statistically supported phylogenetic analyses (Felsenstein, 1989). We cannot, therefore, formally assign confidence limits to the distance tree topology. It should be noted though, that except for two sequences with between 80 and 90% similarity to members of a similarity group, nearest neighbours to similarity groups were all less than 80% similar to any members of that group. However many times the tree is redrawn, therefore, members of the same similarity group will always cluster together. Furthermore, were the similarity groups to have occurred by random chance (in the occurrence of insertions, deletions and substitutions), variation amongst paralogues of the same similarity group should have either occurred randomly along the length of the gene, or at least followed the general pattern of variation amongst p44 paralogues in having more variable central regions with more conserved flanking regions. The precise opposite was observed: paralogues within ‘similarity groups’ had significantly higher similarity in their central domains compared to that occurring in the flanking regions. Paralogues of similarity groups were, therefore, different to other paralogues in both the degree of similarity and the pattern of variation in the different gene regions. The central hydrophilic domains are those most likely to form ligands with host cells (Lin et al., 2002; Zhi et al., 2002a), and high similarity in these regions suggests functional similarity amongst the members of each similarity group. This is more in line with the observation that at least p44 paralogues are involved in specific interactions with host or vector cells that are important for survival of *A. phagocytophilum* (Park et al., 2003). Such paralogues would be more likely to be conserved if the bacterium were to remain infective for the same host species.

The predominantly high degree of variation we observed amongst p44 sequences from UK sheep-derived isolates underlines the potential for phenotypic variation of *A. phagocytophilum* even within foci of infection. The existence of ‘similarity groups’, however, suggests that a few paralogues are more highly conserved amongst isolates, which, being potentially important for adherence to host cells, may provide insight into the host or vector species that an isolate is capable of infecting. In the sample of sequences available to us, ‘similarity groups’ were nearly always country specific (even though paralogues from UK isolates were amplified using primers known to amplify paralogues in USA isolates) and more likely to occur amongst USA rather than UK isolates. At most, these findings may suggest that indeed some specific p44 paralogues are conserved and adapted to different reservoir hosts or vectors in different geographical regions, with corresponding differences in infectivity for different non-reservoir vertebrate species such as humans. Further studies are required, however, to increase the sample size of paralogues and hosts of origin from both Europe and North America to confirm this. At least, our findings suggest that the USA isolates investigated to date (mostly from humans) contained a lower diversity of p44 paralogue variants than did UK isolates from sheep, i.e. the potential for phenotypic variation of *A. phagocytophilum* isolates infecting humans may be considerably less than that occurring in nature. We did find one ‘similarity group’ that comprised paralogues from ruminant UK and human USA isolates, which was the ‘p44-18 group’. It has been hypothesized that the p44-18 paralogue may have a somewhat specific role in the initial phase of infection of previously naïve animals and humans (Zhi et al., 2002b).

In summary, partial 16S rRNA gene sequences did not, for the most part, discriminate amongst UK sheep isolates and isolates from other species (including humans suffering clinical HGE) in other countries. Partial p44 paralogues were amplified from all UK isolates and these had similarities in their DNA and putative peptide sequences, with p44 paralogues amplified from isolates from humans and horses in the USA. A smaller number of paralogues formed ‘similarity groups’, which, in the sample available to us, were more common amongst USA isolates and rarely comprised paralogues from isolates from both the UK and the USA. As paralogues of similarity groups were particularly similar in gene regions likely to interact with the host, they may be useful targets for investigations and identification of any adaptations of *A. phagocytophilum* isolates to different vertebrate host species. Further study of similarity groups of p44 paralogues may, therefore, be fruitful in predicting the infectivity for humans, of different field isolates of *A. phagocytophilum*.
ACKNOWLEDGEMENTS

This study was funded by the Biotechnology and Biological Sciences Research Council of the United Kingdom. All animals used in this study were used in accordance with the relevant Home Office licences.

REFERENCES


