Structural analysis of a *p44/msp2* expression site of *Anaplasma phagocytophilum* in naturally infected ticks in Japan

Wuritu,¹ Yutaka Ozawa,¹ Gaowa,¹ Fumihiko Kawamori,^{1,2}
Takashi Masuda,² Toshiyuki Masuzawa,³ Hiromi Fujita⁴ and Norio Ohashi¹

Correspondence

Norio Ohashi

ohashi@u-shizuoka-ken.ac.jp

¹Laboratory of Microbiology, Department of Food and Nutritional Sciences, Graduate School of Nutritional and Environmental Sciences, University of Shizuoka and Global COE Program, Shizuoka 422-8526, Japan

²Department of Microbiology, Shizuoka Institute of Environment and Hygiene, Shizuoka, Japan

³Laboratory of Microbiology and Immunology, Faculty of Pharmacy, Chiba Institute of Science, Choshi 288-0025, Japan

⁴Ohara Research Laboratory, Ohara General Hospital, Fukushima 960-0195, Japan

Anaplasma phagocytophilum, an agent of human granulocytic anaplasmosis, infects neutrophils and causes an emerging tickborne febrile disease. The genome of this bacterium contains a large number of *p44/msp2*-related genes encoding 44 kDa major outer-membrane proteins, and it is known that a specific *p44/msp2* gene is predominantly transcribed from a single expression locus. This study successfully characterized the genomic expression site for *p44/msp2* (3.8 kb) in uncultured *A. phagocytophilum* from *Ixodes persulcatus* ticks inhabiting a northern part of Japan. Comparative analysis of the sequences revealed that the structures of the expression sites in Japanese *A. phagocytophilum* were similar to those of US strains from human patients and European strains from a dog and sheep, but *omp-1N* (upstream from *p44/msp2*) and a truncated *recA* (downstream from *p44/msp2*) in the *p44/msp2* expression site seemed to share similarities with those of US and European strains. The central hypervariable region sequences of Japanese *p44/msp2* were found to be quite diverse (24.4–100% amino acid similarities) and distinct from their closest relatives from US human patients or animal host origins (56.3–97.6% amino acid similarities) with some exceptions. Thus, this study provides significant information about the molecular characteristics of *A. phagocytophilum* in East Asia, as well as the global diversity of *p44/msp2*.

Received 9 April 2009

Accepted 20 August 2009

INTRODUCTION

Anaplasma phagocytophilum, an obligate and Gram-negative intracellular bacterium, infects neutrophils and causes a febrile systemic illness, granulocytic anaplasmosis, in dogs, cattle, horses and sheep, as well as in humans (Dumler *et al.*, 2005). In nature, *A. phagocytophilum* is maintained in a tick–mammal cycle. The arthropod vectors for the transmission of this agent are primarily *Ixodes* ticks, such as *Ixodes scapularis* and *Ixodes pacificus* in the USA, *Ixodes ricinus* in Europe, and *Ixodes persulcatus* and *Ixodes ovatus* in East Asia including Japan (Dumler *et al.*, 2005; Ohashi *et al.*, 2005).

The GenBank/EMBL/DDBJ accession numbers for the *A. phagocytophilum* *p44/msp2* gene sequences determined in this study are FJ600552–FJ600615. The sequences of *p44/msp2* expression sites in Tick-162-18ES-Iwate-IP and Tick-176-5ES-Iwate-IP are FJ600595 and FJ600601, respectively.

The *p44/msp2* multigene family comprising paralogous genes encoding the 44 kDa major outer-membrane proteins or major surface protein 2 (Barbet *et al.*, 2003; Zhi *et al.*, 1998, 1999, 2002) is thought to be responsible for antigenic variation of *A. phagocytophilum* due to gene conversion (Granquist *et al.*, 2008). Protein structures encoded by this multigene family consist of a single and central hypervariable region flanked by N- and C-terminal conserved regions (Lin *et al.*, 2002). The complete genome sequence of the *A. phagocytophilum* HZ strain found in the USA (Hotopp *et al.*, 2006) revealed 113 *p44/msp2* loci including truncated or short 5' or 3' gene fragments that are widely distributed in the genome. These loci might function as donors, not donor repertoire genes, through (not due to) homologous recombination via the RecF pathway at a single expression locus (Lin *et al.*, 2003, 2006). By this mechanism, *A. phagocytophilum* could express a variety of specific 44 kDa proteins on the bacterial surface

to avoid the host immune defence system (Barbet *et al.*, 2003; Wang *et al.*, 2004, 2007).

In Japan, we previously identified *Ixodes* ticks naturally infected with *A. phagocytophilum* (Ohashi *et al.*, 2005) and characterized the hypervariable regions of *p44/msp2*-related genes (Wuritu *et al.*, 2009). However, as *A. phagocytophilum* has not been isolated in Japan, it is difficult to analyse in detail the molecular and biological features of Japanese *A. phagocytophilum*. In this study, we used an isothermal strand-displacement amplification method to obtain total DNA sufficient for molecular analysis from naturally infected ticks. In this way, we eventually succeeded in the characterization of a genomic expression site for *p44/msp2* in uncultured *A. phagocytophilum* from *Ixodes* ticks.

METHODS

Identification of *A. phagocytophilum*-infected ticks. In 2006, a total of 130 adult and nymph unfed ticks (32 *I. persulcatus*, 96 *I.*

ovatus and two *Haemaphysalis japonica*) were collected by flagging vegetation on wooded areas or fields in the northern parts of Honshu Island including Aomori and Iwate prefectures in Japan. The ticks were dissected and their salivary glands were obtained. Total DNA was extracted from the salivary glands of individual ticks using a QIAamp DNA Mini kit (Qiagen). To identify *A. phagocytophilum*-infected ticks, the hypervariable regions of *p44/msp2* genes specific for *A. phagocytophilum* were amplified from the total DNA prepared from the salivary glands of ticks by nested PCR. The primer pairs used for this purpose were p3726 and p4257 for the first step and p3761 and p4183 for the second step (Table 1; producing the 400 bp fragment A shown in Fig. 1), as described previously (Masuzawa *et al.*, 2008; Ohashi *et al.*, 2005; Wuritu *et al.*, 2009). The PCR products were gel-purified and cloned into a pCR2.1 vector using a TA Cloning kit (Invitrogen). *Escherichia coli* DH5 α (Toyobo) was transformed with the recombinant plasmids. The recombinant clones obtained were selected randomly and the insert DNA of each clone was sequenced.

Analysis of the *p44/msp2* expression site of *A. phagocytophilum* in salivary glands of naturally infected ticks. To yield sufficient DNA for molecular analysis, we first amplified total DNA prepared from the salivary glands of ticks that were PCR-positive for

Table 1. PCR primers used for analysis of a *p44/msp2* expression site of Japanese *A. phagocytophilum* in salivary glands of ticks

Primer name	Sequence (5′→3′)	Length (bp)	Direction	Primer position (bp)*	Amplified fragment†	Product length (bp)	Reference
Amplification of <i>p44/msp2</i> hypervariable region							
p3726F	GCTAAGGAGTTAGCTTATGA	20	Forward	1756–1775	A	536	Ohashi <i>et al.</i> (2005)
p4257R	AGAAGATCATAACAAGCATTG	21	Reverse	2291–2271		Ohashi <i>et al.</i> (2005)	
p3761F	CTGCTCT(T/G)GCCAA(A/G)ACCTC	19	Forward	1808–1826		411	Ohashi <i>et al.</i> (2005)
p4183R	CAATAGT(C/T)TTAGCTAGTAACC	21	Reverse	2218–2198		Ohashi <i>et al.</i> (2005)	
Amplification of a <i>p44/msp2</i> expression site							
omp-1X-F1	ATATCCTGATGCTCACCTATC	21	Forward	−21 to −1	E	458	This study
omp-1X-F2	ATGTGGTGTTCGACTAACACT	21	Forward	22–42		416	This study
omp-NN-R1	TACACTGCCAGTTATACCCTT	21	Reverse	437–417			This study
AB1207F	GGGAGTGCTCTGGTTAGATTTAGG	24	Forward	232–255	D	594	Barbet <i>et al.</i> (2006)
ompN-R1	AGGTATAGATCCCATTCTTCTC	22	Reverse	825–804			This study
AB1207Fs	AGTGCTCTGGTTAGATTTAGG	21	Forward	235–255		466	This study
ompN-R2	TTTCCGGCAGTGAACGCGAAGTG	23	Reverse	700–678			This study
ompNc-F1	GGACTGAGTGGTTCITTCGG	20	Forward	564–583	C	1156	This study
p44N-R1	ACCACTATCTCTAATACCCTT	21	Reverse	1719–1699			This study
ompNc-F2	GGAATGCGGTGCACTTCGC	19	Forward	667–685		960	This study
p44N-R2	CATAGCTACAAGCATGTTGTC	21	Reverse	1626–1606			This study
ompNc-F3	GGTGTTGACACTCTGCTTTC	20	Forward	1014–1033		504	This study
p44N-R3	ACTGCCTTAGTCTCTCCGTT	20	Reverse	1517–1498			This study
p44LA-F1	TGGATTACAGTCCAGCGTTTAGCAA-GATAAGAG	33	Forward	1445–1477	B	1588	This study
p44LA-R1	CACACTCAACAGCTTCTACAGGGCT-ATTCTG	31	Reverse	3032–3002			This study
p44C-F1	CGTCTTGATAGTACTAGTC	22	Forward	2500–2521	F	745	This study
ValS-R2	ATCCGCAAATC(A/T)TAGAGGATAT	22	Reverse	3244–3223			This study
ValSC-F1	TCTTCTCCCTGATCTTGTTCT	21	Forward	3169–3189	G	644	This study
ValSC-F2	AGTTATCTTTCCAATCTTC	19	Forward	3192–3210		621	This study
ValSM-R1	ACAGGTAAGCGTCTAGTTACA	21	Reverse	3812–3792			This study

*Nucleotide positions based on 3791 bp sequence (GenBank accession no. FJ600595 for Tick-162-18ES-Iwate-Ip) of a *p44/msp2* expression site in Japanese *A. phagocytophilum*.

†Amplified DNAs corresponding to fragments A–G in Fig. 1.

Fig. 1. Schematic representation of gene organization of a genomic expression site for *p44/msp2* in Japanese *A. phagocytophilum*. Fragments A–G, which were amplified from the tick samples in this study, are indicated by thick lines with facing arrowheads and are also indicated in Table 1. The central hypervariable region flanked by 5' and 3' conserved regions in *p44/msp2* is indicated by var. Potential promoter regions for *p44/msp2* transcription are shown as open boxes with arrows for regions a and b, as reported by Barbet *et al.* (2005) and Wang *et al.* (2007), respectively.

the hypervariable region sequences of *A. phagocytophilum* as described above, by isothermal strand-displacement amplification using an illustra GenomiPhi V2 DNA amplification kit (GE Healthcare). This procedure with GenomiPhi V2 can generate microgram quantities of the salivary gland DNA from nanogram amounts of starting material in only a few hours. For an initial approach, forward and reverse primers (p44LA-F1 and p44LA-R1; Table 1) were designed based on the conserved sequences among five strains of *A. phagocytophilum* including three strains from US human patients (GenBank accession nos NC_007797, AY164490 and AY137510), a strain from Norwegian sheep (norsheep, GenBank accession no. DQ519565) and a strain from a Swedish dog (swdog, GenBank accession no. DQ519566) that have been published previously (Barbet *et al.*, 2003, 2006; Lin *et al.*, 2003). This primer pair was expected to amplify a 1.6 kb DNA fragment (corresponding to fragment B in Fig. 1) containing two tandemly arranged genes consisting of a 5'-truncated *p44/msp2* and a 3'-truncated *recA*-like gene at a *p44/msp2* expression site (Fig. 1). PCR was conducted using LA *Taq* polymerase (TaKaRa) according to the manufacturer's instructions. The amplicons obtained were gel-purified and TA-cloned as described above, and recombinant clones were selected randomly and sequenced. To extend both 5' and 3' directions of the 1.6 kb DNA, we designed additional primers based on conserved sequences among the strains from US human patients, Norwegian sheep and the Swedish dog as described above (Table 1). By using such primers, DNA fragments corresponding to fragments C–G in Fig. 1 were amplified and directly sequenced.

Sequence analysis. Nucleotide sequences obtained in this study were assembled and translated into protein sequences using the Sequencher program (Gene Codes). Homology searches were performed using BLASTN or BLASTP (<http://blast.ncbi.nlm.nih.gov/Blast.cgi>) and the deduced amino acid sequences were aligned using CLUSTAL_X software (Thompson *et al.*, 1997). A phylogram was constructed using the neighbour-joining method (NJplot software: Saitou & Nei, 1987) and bootstrap values were obtained with 1000 resamplings.

RESULTS AND DISCUSSION

We first surveyed ticks infected with *A. phagocytophilum* by nested PCR targeting the hypervariable regions of *p44/msp2*-related genes. Using this method, we detected a 400 bp DNA fragment corresponding to fragment A as shown in Fig. 1 from four *I. persulcatus* (Tick-60, Tick-162, Tick-176 and Tick-199) and two *I. ovatus* (Tick-70 and Tick-71) in 130 unfed ticks that were collected. The prevalence was thus estimated to be 4.6 % (6/130). We next

examined a DNA region of approximately 1.6 kb (corresponding to fragment B in Fig. 1) containing a 5'-truncated *p44/msp2* and a 3'-truncated *recA* gene in the GenomiPhi-amplified DNAs from six *p44/msp2*-positive ticks by a single PCR with primers p44LA-F1 and p44LA-R1 (Table 1) to characterize the genomic expression site for *p44/msp2*. Using this method, we successfully amplified the corresponding fragment from two *I. persulcatus* ticks (Tick-162 and Tick-176), but could not detect the fragment from the four remaining *I. persulcatus* and *I. ovatus* ticks. To attempt further amplification of the 1.6 kb fragment B from these four ticks, a nested PCR with several primers that were additionally designed (not shown in Table 1) was conducted. However, none of the four infected ticks was PCR-positive, although Tick-162 and Tick-176 were also positive in this nested PCR. The reason why this larger-sized fragment could not be detected in all six infected ticks might be due to differences in infectivity of *A. phagocytophilum* in respective ticks. Indeed, in the survey of infected ticks described above, the hypervariable region for *p44/msp2* could be detected in Tick-162 and Tick-176 by the first-step PCR, but was detected in the other four only by nested PCR. This suggests that the ticks (except for Tick-162 and Tick-176) might be poorly infected and that the genomic DNA of *A. phagocytophilum*, essential for the amplification of this larger-sized fragment, could not be obtained from these four ticks even using template DNA amplified with GenomiPhi. To extend the 5' and 3' regions of fragment B obtained from Tick-162 and Tick-176, we amplified and directly sequenced the DNA corresponding to fragments C–G (Fig. 1) using the primer pairs shown in Table 1. As fragment B includes multiple *p44/msp2* genes, each amplicon was cloned and the respective recombinant clones were sequenced. By sequence analysis, we determined the 3791 and 3782 kb sequences of the *p44/msp2* expression sites from uncultured *A. phagocytophilum* in Japanese ticks. The expression sites consisted of *omp-1X*, *omp-1N* (a gene encoding outer-membrane protein 1N originally identified in *Ehrlichia chaffeensis*), *p44/msp2*, a truncated *recA* and *valS* (Fig. 1), which are similar to those of the strains from several mammalian host species such as humans, dog and sheep in the USA and Europe (Barbet *et*

al., 2003, 2006; Lin *et al.*, 2003), suggesting that the gene organization of *p44/msp2* expression sites may be preserved worldwide.

As shown in Fig. 2(a), in a 274 aa sequence alignment of *omp-1N* located upstream from *p44/msp2*, the sequences from the two ticks (Tick-162 and Tick 176) were identical to each other but distinct from those of the strains from a US human patient (ushuman, GenBank accession no. NC_007797), a Swedish dog (swdog, GenBank accession no. DQ519566) and a Norwegian sheep (norsheep, GenBank accession no. DQ519565) with 82–92 % amino acid similarities. The closest relative of *omp-1N* from Japanese ticks was that from the Swedish dog (92 %), whereas the 87 aa sequences of the truncated *recA* from the two ticks were identical to each other and also to that of ushuman; the lowest similarity was 97 % with that of norsheep (data not shown), suggesting that the sequence similarities of *p44/msp2* expression sites appear to be shared worldwide. In Fig. 2(b), amino acid sequence alignment of *p44/msp2* within the expression site revealed that the protein structures such as the central hypervariable region flanked by the N- and C-terminal conserved regions are considerably preserved worldwide. However, some exceptions were observed with a few changes among *A. phagocytophilum* strains from different host species and/or different geographical locations. For example, the N-terminal sequence MKRGKII in the human, bear and wood rat strains from the USA, and sequence MKRGRLI in both the Swedish dog and Norwegian sheep strains, appeared as MRKGKIVL in Japanese *A. phagocytophilum*-infected ticks. Additionally, the end of the N-terminal conserved region sequence LAKTSGKD (indicated by a bar above the sequence in Fig. 2b) and the conserved ‘signature’ residues KVC, C, F and NWPT within the hypervariable regions (indicated by asterisks above the sequence in Fig. 2b), which have been suggested previously by Barbet *et al.* (2006), had some conservative replacements of amino acid residues regarding charge or size that probably form the necessary framework for correct folding of the protein on the surface of the bacterium. To understand the relationship among *p44/msp2* clones from the six naturally infected ticks, phylogenetic analysis was performed based on the hypervariable region sequences (104–133 aa) (Fig. 3). In this tree, the 64 clone sequences obtained were quite diverse (24.4–100 % amino acid similarities) and distinct from the closest relatives (56.3–97.6 % amino acid similarities; shown in italic in Fig. 3), except for the three relatives P44-15b-HZ-US, P44-1-HZ-US and P44-5-HZ-US from US human patients or animal host origins.

Zhi *et al.* (2002) showed previously that only two *p44/msp2* transcript species were detected in the salivary glands of experimentally infected *I. scapularis* ticks, although a variety of *p44/msp2* transcripts emerged in experimentally infected mammalian hosts after tick transmission, suggesting that specific *p44/msp2* transcript species might be expressed for adaptation of *A. phagocytophilum* in tick host environments. However, in this study, 19 different *p44/*

msp2 sequences within the expression site (shown in bold in Fig. 3; 26.6–100 % amino acid similarities) were identified in two *I. persulcatus* ticks, suggesting a remarkable diversity of *p44/msp2* expression in ticks. This discrepancy between the previous study (Zhi *et al.*, 2002) and this study may depend on the different samples of ticks tested, i.e. experimentally and naturally infected. For the experimental transmission study by Zhi *et al.* (2002), a single strain of *A. phagocytophilum* isolated from a human patient was used, but in nature numerous *A. phagocytophilum* genotypic variants seem to be present (Ladbury *et al.*, 2008; Massung *et al.*, 2005, 2006; Wuritu *et al.*, 2009), probably in naturally infected ticks as well as in wild mammalian reservoirs. In fact, Ladbury *et al.* (2008) reported that multiple strains of *A. phagocytophilum* were probably circulating in lambs within a naturally infected sheep flock. Therefore, the variation in *p44/msp2* within the expression site observed in this study might be generated by the populations of different genotypic variants of *A. phagocytophilum* in naturally infected ticks. Recently, we characterized *p44/msp2* hypervariable region sequences obtained from *A. phagocytophilum* in *Ixodes* ticks collected in Shizuoka and Yamanashi prefectures, in the central regions of Japan (Wuritu *et al.*, 2009), which are geographically different from the regions where the ticks were collected in this study (northern parts of Japan). We compared a group of 64 hypervariable region sequences from this study with another group of 80 sequences from our previous study (Wuritu *et al.*, 2009) and found that there were sequence similarities between the two groups (24.4–99.2 %), except for two identical sequences. This suggests that the distribution of *A. phagocytophilum* genetic variants is dependent on these two different geographical locations. Barbet *et al.* (2003) and Lin *et al.* (2003) first reported a single genomic *p44/msp2* expression locus consisting of transcriptional regulator 1 (*tr1*), *omp-1X*, *omp-1N*, *p44/msp2* and a truncated *recA* pseudogene in *A. phagocytophilum* from human origins. Potential active promoter regions of the expression site, which are located upstream from *omp-1N* (Barbet *et al.*, 2005) for the polycistronic transcript (region a in Fig. 1) and upstream from *p44/msp2* (Wang *et al.*, 2007) for the monocistronic transcript (region b in Fig. 1), have been identified previously. In these promoters, Wang *et al.* (2007) suggested that the transcription of *p44/msp2* in tick salivary glands and mammalian hosts seems to be regulated selectively by these two different promoters: region a for ticks and region b for mammalian hosts. In Japanese *A. phagocytophilum*, the putative promoter region sequences upstream from *omp-1N* (GenBank accession nos FJ600595 and FJ600601; region a in Fig. 1) were identical to previously identified sequences (TAACTT and TTGACT in the –10 and –35 regions, respectively) in three strains from US human patients (GenBank accession nos NC_007797, AY164490 and AY137510), Norwegian sheep (GenBank accession no. DQ519565) and a Swedish dog (GenBank accession no. DQ519566), suggesting worldwide preservation and probably similar promoter activities

(a)	1	MVFRGLRCLFFTCCLLVSSCLVERGEADGFFYASLGWGESVADFGVEHLKV	Tick-162-18ES-Iwate-lp
	1	MVFRGLRCLFFTCCLLVSSCLVERGEADGFFYASLGWGESVADFGVEHLKV	Tick-176-5ES-Iwate-lp
	1	MVFRGLRCLFFTCCLLVSSCLVERGEADGFFYASLGWGESVADFGVEHLKV	ushuman
	1	MVFRGLRCLFFTCCLLVSSCLVERGEADGFFYASLGWGESVADFGVEHLKV	swdog
	1	MVFRGLRCLFFTCCLLVSSCLVERGEADGFFYASLGWGESVADFGVEHLKV	norsheep
	51	KGITGSGVMGPRKSELDAVKYSRNTFSKGI GGVGSGEF GDKYRPVYTEDFG	Tick-162-18ES-Iwate-lp
	51	KGITGSGVMGPRKSELDAVKYSRNTFSKGI GGVGSGEF GDKYRPVYTEDFG	Tick-176-5ES-Iwate-lp
	51	KGITGSGVMGPRKSELDAVKYGHDTFSKGI GGVGSGEF GDKYRPVYTEDFG	ushuman
	51	KGITGSGVMGPRKSELDAVKYGHDTFSKGI GGVGSGEF GDKYRPVYTEDFG	swdog
	51	KGITGSGVMGPRKSELDAVKYGHDTFSKGI GGVGSGEF GDKYRPVYTEDFG	norsheep
	101	LSGSFGYRFGNLGIEIEGSEQEFHPNTAGYKVEENAVHFAFTAGKRGVNE	Tick-162-18ES-Iwate-lp
	101	LSGSFGYRFGNLGIEIEGSEQEFHPNTAGYKVEENAVHFAFTAGKRGVNE	Tick-176-5ES-Iwate-lp
	101	LSGSFGYRFGNLGIEIEGSEQEFHPNTAGYKVEENAVHFAFTAGKRGVNE	ushuman
	101	LSGSFGYRFGNLGIEIEGSEQEFHPNTAGYKVEENAVHFAFTAGKRGVNE	swdog
	101	LSGSFGYRFGNLGIEIEGSEQEFHPNTAGYKVEENAVHFAFTAGKRGVNE	norsheep
	151	KVSGFVGGENPGMHLGLVLLNGCYSGRLRLDEKNGIYTCLGLGLAAADYLG	Tick-162-18ES-Iwate-lp
	151	KVSGFVGGENPGMHLGLVLLNGCYSGRLRLDEKNGIYTCLGLGLAAADYLG	Tick-176-5ES-Iwate-lp
	151	KVSGFVGGENPGMHLGLVLLNGCYSGRLRLDEKNGIYTCLGLGLAAADYLG	ushuman
	151	KVSGFVGGENPGMHLGLVLLNGCYSGRLRLDEKNGIYTCLGLGLAAADYLG	swdog
	151	KVSGFVGGENPGMHLGLVLLNGCYSGRLRLDEKNGIYTCLGLGLAAADYLG	norsheep
	201	LGRRLRALWDAAKIGVELQFTKNLAFGEVYRGGFVLLGRVVRVFSVAALEGE	Tick-162-18ES-Iwate-lp
	201	LGRRLRALWDAAKIGVELQFTKNLAFGEVYRGGFVLLGRVVRVFSVAALEGE	Tick-176-5ES-Iwate-lp
	201	LGRRLRALWDAAKIGVELQFTKNLAFGEVYRGGFVLLGRVVRVFSVAALEGE	ushuman
	201	LGRRLRALWDAAKIGVELQFTKNLAFGEVYRGGFVLLGRVVRVFSVAALEGE	swdog
	201	LGRRLRALWDAAKIGVELQFTKNLAFGEVYRGGFVLLGRVVRVFSVAALEGE	norsheep
	251	VDTLLSSGMDVSVVGNACGIRYLF	Tick-162-18ES-Iwate-lp
	251	VDTLLSSGMDVSVVGNACGIRYLF	Tick-176-5ES-Iwate-lp
	251	VDTLLSSGMDVSVVGNACGIRYLF	ushuman
	251	VDTLLSSGMDVSVVGNACGIRYLF	swdog
	251	VDTLLSSGMDVSVVGNACGIRYLF	norsheep
(b)	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	Tick-162-18ES-Iwate-lp
	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	Tick-176-5ES-Iwate-lp
	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	ushuman
	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	uswat
	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	usbear
	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	swdog
	1	MRKGGKVLGSVMMMSMAIVMAGNDVRAHDDVSALDTGGAGYFYVGLDYSPA	norsheep
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	Tick-162-18ES-Iwate-lp
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	Tick-176-5ES-Iwate-lp
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	ushuman
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	uswat
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	usbear
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	swdog
	51	FSKIRDFSIRESNGETKAVYPYLKDGKSVKLESHKFDWNTPDPRIGFKDN	norsheep
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	Tick-162-18ES-Iwate-lp
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	Tick-176-5ES-Iwate-lp
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	ushuman
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	uswat
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	usbear
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	swdog
	101	MLVAMEGSVGYGIGGARVELEIGYERFKTKGIRDSSGSKEDAEADTVYLLAK	norsheep
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	Tick-162-18ES-Iwate-lp
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	Tick-176-5ES-Iwate-lp
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	ushuman
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	uswat
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	usbear
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	swdog
	151	ELAYDVVTGQTDNLAAALAKTSGKDIVQFAKAVEISHPHIDGKVCSSGSHA	norsheep
	201	GSDGN--SYGVYAEIKAGTGNAGTAFAGGVGRSSGGNQAA--TTPQVLK	Tick-162-18ES-Iwate-lp
	201	ALVTSKGTAYN--AD--PGTTDSNTAQCSGLGGTGAATGPK--S--FS	Tick-176-5ES-Iwate-lp
	201	GGDTN--RFAKYIVAGAGD--SSNAGTSLGCKNQKSSDITDGVKADALH	ushuman
	201	DRDKSHQAGTATFG--DKAQATVAAQCSGFAT--KKEG--LTF	uswat
	201	AVGIAAQKKEYARVP--DSSSAGKTAQCSGFKNKSDAKNEH--KFLS	usbear
	201	ALSESASAAADYASDFS--SGSTTAKTAQCSGFKNKSGHTTEK--GGLN	swdog
	201	SSSKNK--TGAYGQYALS--NN--SADHTVALGGA--GHASNGSS--SPQYLK	norsheep
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	Tick-162-18ES-Iwate-lp
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	Tick-176-5ES-Iwate-lp
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	ushuman
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	uswat
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	usbear
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	swdog
	245	DFVRVTLKGDGSKNWPTSTGTGAPT--K--PNDNAEAVAKDLVA--LNS	norsheep
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	Tick-162-18ES-Iwate-lp
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	Tick-176-5ES-Iwate-lp
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	ushuman
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	uswat
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	usbear
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	swdog
	289	DEKTI VAGLLAKTIEGGGEVVEIRAVSSTSVMVNACYDLLSEGLGVVPYAC	norsheep
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	Tick-162-18ES-Iwate-lp
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	Tick-176-5ES-Iwate-lp
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	ushuman
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	uswat
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	usbear
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	swdog
	339	VGLGGNFVGVVDGHITPKLAYRLKAGLSYQLSPEISAFAGGFYHRVVDGDG	norsheep
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	Tick-162-18ES-Iwate-lp
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	Tick-176-5ES-Iwate-lp
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	ushuman
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	uswat
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	usbear
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	swdog
	389	VYDDLPAQRLVDDTSPAGRTKDTAIAINFSMAYVGGGEFGVRFAP	norsheep

Fig. 2. Variation in amino acid sequences of *omp-1N* (a) and *p44/msp2* (b) within the expression site in *A. phagocytophilum* among different host species (Tick-162-18ES-lwate-lp and Tick-176-5ES-lwate-lp for Japanese *I. persulcatus* ticks, GenBank accession nos FJ600595 and FJ600601, respectively; ushuman for US human patient, NC_007797; uswrat for US wood rat, DQ510570; usbear for US bear, DQ519567; swdog for Swedish dog, DQ519568; and norsheep for Norwegian sheep, DQ519569). In (b), the bars above the sequence indicate the end of the N-terminal conserved region sequence and the beginning of the C-terminal conserved region sequence. Asterisks indicate the conserved 'signature' residues KVC, C, F and NWPT with some residue replacements within the hypervariable regions of *p44/msp2*, which has been suggested by Barbet *et al.* (2006).

among them in ticks. However, the sequences of promoters located upstream from *p44/msp2* (region b in Fig. 1) were quite diverse among the above different host species (data not shown), suggesting distinctive promoter activities for transcriptional regulation in various mammalian host environments.

As described above, our data present here, for the first time, characterization of *p44/msp2* expression sites from uncultured *A. phagocytophilum* in naturally infected ticks. This study provides significant information about the molecular characteristics of *A. phagocytophilum* in East Asia.

Fig. 3. Relationships among the hypervariable region sequences of *p44/msp2* species from *A. phagocytophilum* in ticks of *I. persulcatus* (lp) and *I. ovatus* (lo) by phylogenetic analysis. A phylogram was constructed by the neighbour-joining method based on amino acid sequences of 104–133 aa of *p44/msp2* clones from *A. phagocytophilum* in six ticks (Tick-60, Tick-70, Tick-71, Tick-162, Tick-176 and Tick-199). The *p44/msp2* clone sequences located at the expression sites are indicated in bold and the closely related *p44/msp2* from US patients, a US wood rat, a Swedish dog and Norwegian sheep are shown in italic. The numbers of *p44/msp2* clones with identical amino acid sequences and/or accession numbers of the sequences are shown in parentheses. Numbers at the nodes in the tree show the bootstrap values with 1000 resamplings. Bar, 5% sequence divergence.

ACKNOWLEDGEMENTS

This work was supported in part by a grant for Research on Emerging and Re-emerging Infectious Diseases from the Japanese Ministry of Health, Labour and Welfare (H18-Shinkou-Ippan-014 and H21-Shinkou-Ippan-006), and a Grant-in-Aid for Scientific Research (C) from the Japan Society for the Promotion of Science (no. 19590451) for N. O.

REFERENCES

- Barbet, A. F., Meeus, P. F., Bélanger, M., Bowie, M. V., Yi, J., Lundgren, A. M., Alleman, A. R., Wong, S. J., Chu, F. K. & other authors (2003). Expression of multiple outer membrane protein sequence variants from a single genomic locus of *Anaplasma phagocytophilum*. *Infect Immun* **71**, 1706–1718.
- Barbet, A. F., Agnes, J. T., Moreland, A. L., Lundgren, A. M., Alleman, A. R., Noh, S. M., Brayton, K. A., Munderloh, U. G. & Palmer, G. H. (2005). Identification of functional promoters in the *msp2* expression loci of *Anaplasma marginale* and *Anaplasma phagocytophilum*. *Gene* **353**, 89–97.
- Barbet, A. F., Lundgren, A. M., Alleman, A. R., Stuen, S., Bjöersdorff, A., Brown, R. N., Drzenovitch, N. L. & Foley, J. E. (2006). Structure of the expression site reveals global diversity in MSP2 (P44) variants in *Anaplasma phagocytophilum*. *Infect Immun* **74**, 6429–6437.
- Dumler, J. S., Choi, K. S., Garcia-Garcia, J. C., Barat, N. S., Scorpio, D. G., Garyu, J. W., Grab, D. J. & Bakken, J. S. (2005). Human granulocytic anaplasmosis and *Anaplasma phagocytophilum*. *Emerg Infect Dis* **11**, 1828–1834.
- Granquist, E. G., Stuen, S., Lundgren, A. M., Bråten, M. & Barbet, A. F. (2008). Outer membrane protein sequence variation in lambs experimentally infected with *Anaplasma phagocytophilum*. *Infect Immun* **76**, 120–126.
- Hotopp, J. C., Lin, M., Madupu, R., Crabtree, J., Angiuoli, S. V., Eisen, J. A., Seshadri, R., Ren, Q., Wu, M. & other authors (2006). Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet* **2**, e21.
- Ladbury, G. A., Stuen, S., Thomas, R., Bown, K. J., Woldehiwet, Z., Granquist, E. G., Bergström, K. & Birtles, R. J. (2008). Dynamic transmission of numerous *Anaplasma phagocytophilum* genotypes among lambs in an infected sheep flock in an area of anaplasmosis endemicity. *J Clin Microbiol* **46**, 1686–1691.
- Lin, Q., Zhi, N., Ohashi, N., Horowitz, H. W., Aguero-Rosenfeld, M. E., Raffalli, J., Wormser, G. P. & Rikihisa, Y. (2002). Analysis of sequences and loci of *p44* homologs expressed by *Anaplasma phagocytophilum* in acutely infected patients. *J Clin Microbiol* **40**, 2981–2988.
- Lin, Q., Rikihisa, Y., Ohashi, N. & Zhi, N. (2003). Mechanisms of variable *p44* expression by *Anaplasma phagocytophilum*. *Infect Immun* **71**, 5650–5661.
- Lin, Q., Zhang, C. & Rikihisa, Y. (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.
- Massung, R. F., Courtney, J. W., Hiratzka, S. L., Pitzer, V. E., Smith, G. & Dryden, R. L. (2005). *Anaplasma phagocytophilum* in white-tailed deer. *Emerg Infect Dis* **11**, 1604–1606.
- Massung, R. F., Mather, T. N. & Levin, M. L. (2006). Reservoir competency of goats for the Ap-variant 1 strain of *Anaplasma phagocytophilum*. *Infect Immun* **74**, 1373–1375.
- Masuzawa, T., Kharitonov, I. G., Okamoto, Y., Fukui, T. & Ohashi, N. (2008). Prevalence of *Anaplasma phagocytophilum* and its coinfection with *Borrelia afzelii* in *Ixodes ricinus* and *Ixodes persulcatus* ticks inhabiting Tver Province (Russia) – a sympatric region for both tick species. *J Med Microbiol* **57**, 986–991.
- Ohashi, N., Inayoshi, M., Kitamura, K., Kawamori, F., Kawaguchi, D., Nishimura, Y., Naitou, H., Hiroi, M. & Masuzawa, T. (2005). *Anaplasma phagocytophilum*-infected ticks, Japan. *Emerg Infect Dis* **11**, 1780–1783.
- Saitou, N. & Nei, M. (1987). The neighbor-joining method: a new method for reconstructing phylogenetic trees. *Mol Biol Evol* **4**, 406–425.
- Thompson, J. D., Gibson, T. J., Plewniak, F., Jeanmougin, F. & Higgins, D. G. (1997). The CLUSTAL_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* **25**, 4876–4882.
- Wang, X., Rikihisa, Y., Lai, T. H., Kumagai, Y., Zhi, N. & Reed, S. M. (2004). Rapid sequential changeover of expressed *p44* genes during the acute phase of *Anaplasma phagocytophilum* infection in horses. *Infect Immun* **72**, 6852–6859.
- Wang, X., Cheng, Z., Zhang, C., Kikuchi, T. & Rikihisa, Y. (2007). *Anaplasma phagocytophilum* *p44* mRNA expression is differentially regulated in mammalian and tick host cells: involvement of the DNA binding protein ApxR. *J Bacteriol* **189**, 8651–8659.
- Wuritu, Gaowa, Kawamori, F., Aochi, M., Masuda, T. & Ohashi, N. (2009). Characterization of *p44/msp2* multigene family of *Anaplasma phagocytophilum* from two different tick species, *Ixodes persulcatus* and *Ixodes ovatus*, in Japan. *Jpn J Infect Dis* **62**, 142–145.
- Zhi, N., Ohashi, N., Rikihisa, Y., Horowitz, H. W., Wormser, G. P. & Hechemy, K. (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.
- Zhi, N., Ohashi, N. & Rikihisa, Y. (1999). Multiple *p44* genes encoding major outer membrane proteins are expressed in the human granulocytic ehrlichiosis agent. *J Biol Chem* **274**, 17828–17836.
- Zhi, N., Ohashi, N., Tajima, T., Mott, J., Stich, R. W., Grover, D., Telford, S. R., III, Lin, Q. & Rikihisa, Y. (2002). Transcript heterogeneity of the *p44* multigene family in a human granulocytic ehrlichiosis agent transmitted by ticks. *Infect Immun* **70**, 1175–1184.