



ELSEVIER

Contents lists available at ScienceDirect

Gene: X

journal homepage: www.journals.elsevier.com/gene-x

Dynamics of repeat-associated plasticity in the *aaap* gene family in *Anaplasma marginale*



Heather M. Fallquist^a, Jin Tao^b, Xiaoya Cheng^a, Sebastian Aguilar Pierlé^a, Shira L. Broschat^{a,b,c}, Kelly A. Brayton^{a,b,c,*}

^a Program in Genomics, Department of Veterinary Microbiology & Pathology, Washington State University, Pullman, WA 99164-7040, USA

^b School of Electrical Engineering and Computer Science, Washington State University, Pullman, WA 99164-2752, USA

^c Paul G. Allen School for Global Animal Health, Washington State University, Pullman, WA 99164-7090, USA

ARTICLE INFO

Keywords:

Appendage

Alp

Genome plasticity

Tick-borne pathogen

ABSTRACT

Anaplasmosis, the most prevalent tick-transmitted disease of cattle, is caused by the rickettsial intracellular parasite *Anaplasma marginale*. The pathogen replicates within a parasitophorous vacuole formed from the invagination of the erythrocyte membrane. Several strains of *A. marginale* form “tails” or “appendages” which are attached to, and extend out from, the cytoplasmic side of the parasitophorous vacuole. Genomic analysis of the parasite antigen distributed along the appendage led to the discovery of the *aaap* (*Anaplasma* appendage associated protein) gene family located within a highly plastic region in the genome. The *aaap* gene family consists of *aaap* and several *alps* (for *aaap*-like proteins), depending on the strain. These genes/proteins are characterized by repeat sequences. To investigate locus plasticity, different versions of the locus were cloned from the same strain as well as from different strains, sequenced and aligned to identify changes. Our findings show that repeat sequences both within and between genes facilitated rearrangement events within the locus. Structural variation of the locus in the St. Maries strain was further investigated during infection of different cellular environments, i.e., bovine erythrocytes and tick cells, with a reduction in subpopulations of the *aaap* locus within the tick as compared to erythrocytes. Interestingly, subpopulations bearing alternative locus structures began to arise again when the pathogen was transferred from the tick environment into a naïve calf. Additionally, the Aaap protein expression profile between blood and tick samples showed a regulatory shift, indicating a host-specific response. Alignment of the protein sequences from different species of *Anaplasma* reveals six similar repeating motifs that appear to be unique to a few species of *Anaplasma*. The role the *aaap* locus may play in the pathogenesis of the bovine host or in tick infection/transmission remains unknown; however, the changes in *aaap* locus subpopulations, locus structure, and protein expression indicate that these genes have a role in strain diversification.

1. Introduction

Anaplasma marginale is the most prevalent tick-transmitted pathogen of cattle and the causative agent of anaplasmosis. Clinical disease manifests with fever, weight loss, lethargy, anemia, and often death in animals over two years old (Kocan et al., 2003; Brayton et al., 2005). Animals that survive acute infection develop a life-long persistent infection and serve as reservoirs for continual transmission by the arthropod vector. Within the bovine host, the parasite resides and replicates in an inclusion body or parasitophorous vacuole which is formed from the invagination of the mature erythrocyte membrane (Stich et al., 1997; Stich et al., 2004). It was observed that infected

erythrocytes stained with Giemsa's stain, Tiosson's fluid, and new methylene blue and viewed under light microscopy revealed the presence of what appeared to be ‘tails,’ ‘comets,’ and ‘bands’ extending out from the cytoplasmic side of the inclusion body membrane (Boynton, 1932; Kreier and Ristic, 1963; Espana and Espana, 1963; Franklin and Redmond, 1958). Electron microscopy and immunocytochemical studies of the inclusion tails showed these structures to be highly ordered bundles consisting of cross-linked F-actin and parasite antigens (Stich et al., 1997; Kocan et al., 1978a; Kocan et al., 1978b). The parasite encoded antigen found within the inclusion appendage was identified as the *Anaplasma marginale* appendage-associated protein (Aaap) (Stich et al., 2004).

Abbreviations: Aaap, *Anaplasma* appendage associated protein; Alp, Aaap-like protein; RPKM, reads per kilobase per million mapped reads

* Corresponding author at: Department of Veterinary Microbiology & Pathology, Washington State University, Pullman, WA 99164-7040, USA.

E-mail address: kbrayton@wsu.edu (K.A. Brayton).

<https://doi.org/10.1016/j.gene.2019.100010>

Received 5 January 2019; Received in revised form 8 February 2019; Accepted 14 February 2019

Available online 17 February 2019

2590-1583/ © 2019 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license

(<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

The complete genome sequence of the *A. marginale* St. Maries strain revealed the presence of additional members in the *aaap* gene family—*alp1* and *alp2* (Aaap-like protein)—which are found arranged in tandem in the *aaap* locus (Brayton et al., 2005). Comparison of the St. Maries and Florida strain genome sequences shows that despite a high degree of overall genome synteny, the Florida strain shows a marked expansion of the locus which consists of a duplicated *aaap* gene and three *aaap*-like genes. Additionally, the *aaap* genes exhibit relatively low levels of sequence identity. Examination of additional strains reveals that this locus is a very dynamic region of the genome (Dark et al., 2009). Southern analysis comparing the *aaap* loci in the St. Maries, Florida, Mississippi, Puerto Rico, South Idaho and Virginia strains shows a high degree of plasticity in this locus as multiple bands of varying sizes are apparent. Interestingly, while all these strains have *aaap* genes, not all express the inclusion appendage: the level of expression of Aaap in the Florida strain is extremely low compared to other strains and this is thought to preclude appendage formation (Stich et al., 2004). The St. Maries, South Idaho and Virginia strains are known to express the appendage, while appendage status has not been evaluated for the other strains in the study (Stich et al., 2004; McGuire et al., 1984).

Sequence analyses and alignments of the encoded Aaap proteins from the St. Maries, Florida, Illinois, and Virginia strains revealed the proteins differed in length and contained multiple imperfect repeat elements centered around the ELKAIADAE motif (Stich et al., 2004). The high degree of sequence polymorphism between these proteins may be due to rearrangement, insertion, and deletion events in the repetitive DNA sequence. While the direct role these repeat elements play in these proteins is unknown, we examined whether these repeat elements mediate locus plasticity. Our approach was to analyze multiple cloned loci for evidence of locus expansion and contraction. Furthermore, we examined locus variation within infected tick embryonic ISE6 cells and determined if there were changes in locus structure when passaged back into the bovine host. Finally, we tested for differential expression of the *aaap* genes through transcriptional and protein analysis in the tick and bovine host environments. These investigations into locus plasticity and structure within *A. marginale* strains are the first steps towards determining the functionality of these genes and their protein products in the bovine host and tick vector.

2. Materials and methods

2.1. Strains and blood preparation

The origin and isolation of the *Anaplasma marginale* strains St. Maries, Virginia, Florida, Puerto Rico, Mississippi, South Idaho, Virginia, EMΦ, and 6DE have been previously reported (McGuire et al., 1984; Eriks et al., 1994; Kuttler and Winward, 1984; Ristic and Carson, 1977; Scoles et al., 2007; Hidalgo et al., 1989; Palmer et al., 2004). Blood collected for DNA and protein isolation was washed 3 to 7 times in PBS at 1500 ×g for 10 min with the removal of the buffy coat after each spin. Washed blood was diluted 1:1 in PBS for storage at −20 °C.

2.2. Southern analysis

Genomic DNA from infected bovine blood and ISE6 cells was isolated using the Genra Puregene Blood Kit (Qiagen). DNA was digested with restriction enzymes *HindIII* and *XbaI* (New England Biolabs) as these enzymes cut in conserved genes flanking the *aaap* locus. DNA was separated on a 0.7% agarose gel and subsequently transferred to a nylon membrane (Roche Applied Science). The blots were pre-hybridized at 42 °C for a minimum of two hours in Dig Easy Hyb buffer (Roche Applied Science). Digoxigenin-labeled probes were synthesized with primers *aaapF* 5'- ATT GTG ACA TAT GGC ACT GTG GG -3' and *aaapR* 5'- GGA CCC CAA GCA TCC AAG AAA C -3' or *msp5F* 5' - CGG CGA GAG GTT TAC CAC TTC C and *msp5R* 5' - GTG CTT GCC GCC

AAA ATC G using the PCR DIG Probe Synthesis Kit (Roche Applied Science) and added to the buffer to incubate at 42 °C overnight. The *msp5* probe was included to control for the amount and integrity of the DNA and for complete digestion by the restriction endonucleases. The membranes were washed 3 times at 65 °C for 15 min in 2× SSC, 0.1% SDS, followed by a final wash in 0.2× SSC, 0.1% SDS. Chemiluminescent detection was performed using CDP-Star as directed (Roche Applied Science).

2.3. Protein preparation and Western analysis

Stored blood was washed 5–7 times with PBS at 30,000 ×g for 25 min to remove hemoglobin. The pellets were resuspended in lysis buffer (50 mM Tris [pH 8.0], 5 mM EDTA, 1% Nonidet P-40) and sonicated for 2 min in 30-s intervals. The samples were boiled in SDS-PAGE buffer for 5 min. Isolation of proteins from ISE6 cells infected with the St. Maries strain and uninfected ISE6 cells was done as previously described (Ramabu et al., 2010). Protein samples were electrophoresed on precast 12.5% Criterion Tris-HCL Gels (Bio-Rad) at 100 V for 3 h. After transfer to nitrocellulose and overnight blocking in I-Block reagent (Applied Biosystems) with 0.5% Tween 20, proteins were detected using the previously described monoclonal antibodies ANAF16C1 at 0.02 μg/ml for *Msp5* and ANAO24D1 at 3 μg/ml for Aaap proteins (Eriks et al., 1994). *Msp5*, a constitutively expressed protein, was detected as a control to demonstrate equal loading in each lane. Antigen binding was detected with a 1:7000 dilution of alkaline phosphatase-conjugated goat anti-mouse secondary antibody from Western-Star chemiluminescent immunoblot detection systems (Applied Biosystems) according to manufacturer's specifications.

2.4. Cloning of the *aaap* locus

DNA isolated from the St. Maries and Virginia strains was digested with *XbaI* and *HindIII* restriction enzymes (New England Biolabs) and separated on a 0.7% agarose gel. Gel slices corresponding to the appropriate size of the *aaap* locus were extracted using the QIAquick Gel Extraction kit (Qiagen). The DNA was ligated with *XbaI*-*HindIII*-digested pBluescript II KS- (Stratagene) vector. Transformed colonies, grown in *E. coli* TOP10 or Stb12 (Invitrogen) cells, were screened for inserts containing the *aaap* locus by membrane hybridization using the Digoxigenin-labeled *aaap* probe. Membrane hybridization was carried out as directed (Roche Applied Science) with denaturation in 0.5 M NaOH, 1.5 M NaCl for 15 min, neutralization in 0.5 M Tris [pH 7.5], 1.5 M NaCl for 15 min, followed by 10 min in 2× SSC. Hybridization was as described for Southern blots. Cloning of the St. Maries *aaap* locus from DNA isolated from infected ISE6 cells was performed via Polymerase Chain Reaction (PCR) using primers 5'-CAG GCC CAA AAT CGC GTC ATC C-3' and 5'-CCC TAG CCC TAT ATC GGT TGC GAA TA-3'. The ends were cut with the restriction enzymes *HindIII* and *XbaI* and ligated into pBluescript II KS-.

2.5. Sequencing of the *aaap* locus

Plasmid DNA isolated from positive transformants was digested with *HindIII* and *XbaI* enzymes to determine insert size. For sequencing, the inserts were subcloned using the restriction enzymes *SphI*, *XhoI*, *ApaI*, *SmaI*, *BamHI*, *ApaLI*, *KpnI*, *XbaI*, and *HindIII* (New England Biolabs) either alone or in combination and the resulting fragments were religated, sequenced, and assembled. Fragments were sequenced on both strands using the BigDye Terminator v3.1 Cycle Sequencing kit (Invitrogen) and an ABI 3130XL automated sequencer. The sequences of three variant loci from the St. Maries strain obtained from blood stage infection have accession numbers GenBank: MK330651 (clone 3–8, 4535 bp), GenBank: MK330652 (clone 3–6, 4646 bp), GenBank: MK330653 (clone 8–1, 6730 bp). The sequence of the St. Maries strain locus obtained from tick cell culture has accession number GenBank:

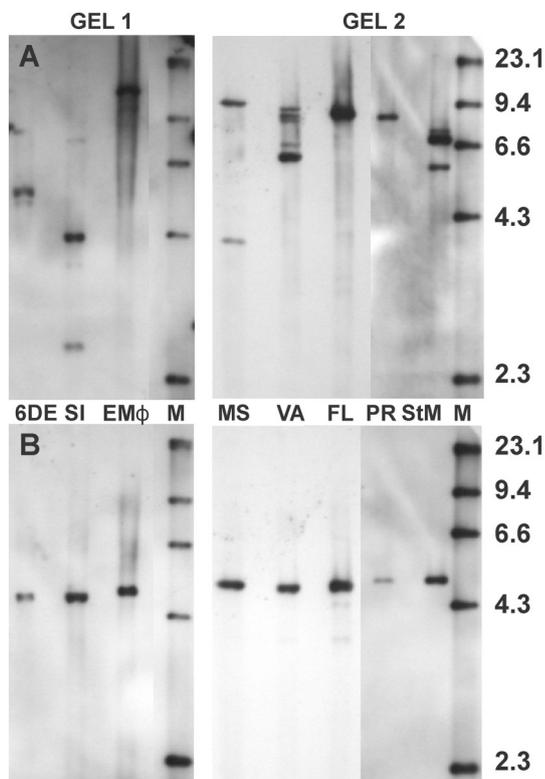


Fig. 1. Southern analysis of *A. marginale* strains. A composite image from two gels containing genomic DNA digested with *Hind*III and *Xba*I is shown, with the images from gel 2 having two different exposure times for better visualization. Strains shown from left to right are 6DE, South Idaho (SI), EM Φ , Mississippi (MS), Virginia (VA), Florida (FL), Puerto Rico (PR), and St. Maries (StM) respectively. M indicates λ HindIII marker. (A) Southern probed with *aaap* (B) Southern probed with the single copy gene *msp5* to allow for visualization of the amount and integrity of the DNA and the complete digestion by the restriction endonucleases.

MK330654 (clone ISE6, 6708 bp). The sequence of the loci obtained from the Virginia strain have accession numbers GenBank: MK330655 (Va small, 6295 bp) and GenBank: MK330656 (Va large, 8136 bp). The original St. Maries and Florida genome versions of the locus are found in GenBank: [CP000030](#) and GenBank: [CP001079](#), respectively.

2.6. Transcriptional mapping of *aaap* locus via RNA-Seq

The accession number for the RNA-Seq portion of this study is GenBank: SRP014580. Briefly, a Holstein calf negative for *A. marginale* by *Msp*5 cELISA and T75 flasks with ISE6 cells were infected with the St. Maries strain of *A. marginale*. *Anaplasma marginale* was maintained in ISE6 cells cultured at 34 °C as previously described (Felsheim et al., 2010; Munderloh et al., 1996; Munderloh et al., 1994). Infection levels in the calf were tracked by analysis of Giemsa stained blood smears to calculate the percentage of parasitized erythrocytes (PPE) or by examination of Giemsa stained cytospin preparations for infected T75 flasks. When the calf reached peak parasitemia and 80% of ISE6 cells were found to be infected, samples were stored in TRIzol (Invitrogen). Total RNA was isolated from *A. marginale*-infected ISE6 cells and blood using TRIzol (Invitrogen) per the manufacturer's directions. Eukaryotic sequences were negatively selected through hybridization using the MICROBEnrich kit (Ambion). The Duplex-Specific thermostable nuclease (DSN) normalization protocol was applied to the generated libraries before Illumina sequencing (Miller et al., 2013). Data was processed using the CLC Genomics Workbench (CLC Bio) as previously described (Pierle et al., 2012). The sequenced *aaap* loci from blood and tick cells (MK330653 [blood] and MK330654 [ISE6 tick cells]) were incorporated into the *A. marginale* chromosome for more accurate mapping of

transcripts for each condition. Fold changes with respect to RPKM (Reads Per Kilobase per Million mapped reads) values were calculated (Mortazavi et al., 2008). RPKM values were adjusted to reduce non-specific counts due to non-specific mapping to repetitive regions.

2.7. Estimating the probability of a motif occurring in a random protein sequence

Alignment of *Aaap*-like protein sequences from different species of *Anaplasma* reveals six similar repeating motifs (ELDAIDA, ALKAIDA, ELKAIDA, ELRAIDA, ELRAIDE and ELRAINA). To eliminate the possibility that a motif is simply a random occurrence of a sequence of amino acids, we can estimate the probability of a motif occurring in a random protein sequence. On September 21, 2017, the total number of bacterial protein sequences in the GenBank NR database (Model XPs excluded) was 56,772,018 with an average length of 329 amino acids. Given our target motif length is 7 amino acids, we calculate the probability of an arbitrary target motif of 7 amino acids occurring in a random sequence of average length. To avoid overlapping motifs, the minimal distance between the starting points of any two motif patterns is 7 amino acids. If we view each motif of 7 amino acids as a sliding window, there are 323 possible windows for an average sequence of 329 amino acids. Moreover, one sequence of average length can have at most 47 (329/7) repetitions of one target motif. Thus, the existence of a target motif in an average sequence can be formulated as selecting n ($n \geq 1$) target cells from an array of length 323 with the distance between any two selected cells being at least 7, with each cell of this array representing a sliding window of 7 amino acids, and we use combinatorics to calculate the probability.

The principle of inclusion-exclusion is a counting method for determining the number of elements in a set by taking the union of all the elements and then systematically decreasing the number by subtracting or adding the intersections of the different sets to adjust for multiple counting. For our case, the number of combinations to give at least one target motif in an average sequence is given by $S = |A_1 \cup A_2 \cup A_3 \cup A_4 \dots \cup A_{323}| = \sum_{1 \leq i \leq 323} |A_i| - \sum_{1 \leq i < j \leq 323} |A_i \cap A_j| + \sum_{1 \leq i < j < k \leq 323} |A_i \cap A_j \cap A_k| - \dots + (-1)^{322} |A_1 \cap A_2 \cap \dots \cap A_{322} \cap A_{323}|$, where A_i is a target cell representing a 7-amino acid motif, the first term $\sum_{1 \leq i \leq 323} |A_i|$ denotes the number of single appearances of a target motif, the second term $\sum_{1 \leq i < j \leq 323} |A_i \cap A_j|$ denotes the number of double appearances of a target motif, and so on. While all the terms are included for completeness, after the 47th term, they all vanish because there can not be > 47 repetitions of the target motif. The final probability of having a motif pattern in a random sequence of average length is estimated by taking the ratio of the number of single occurrences of the motif and the total number of amino acid combinations in an average protein sequence which yields $S / (20^{329}) = 3.2594E - 406$. It should be noted that we did not restrict the amino acid composition in this estimate which would reduce it to an even smaller value. On the other hand, we considered all combinations of amino acids to be of equal likelihood which is not the case, and this would increase the value. Even so, the number would still be very small, and given the number of motif repetitions in *Anaplasma*, it is clearly not a random phenomenon.

3. Results

3.1. Plasticity of the *aaap* locus

Initial analysis of the *aaap* locus shows the inherent plasticity of this locus among *A. marginale* strains (Dark et al., 2009). To further investigate the plasticity of this locus, Southern analysis of eight *A. marginale* strains (Fig. 1A) was performed. The array of band sizes observed across the various strains shows the marked variability of the *aaap* locus between strains. Some strains appeared to contain a single *aaap* locus size such as the Florida, EM Φ and Puerto Rico strains, whereas in other strains, including Virginia, St. Maries, 6DE, Mississippi, and South Idaho, several bands were apparent. This indicates

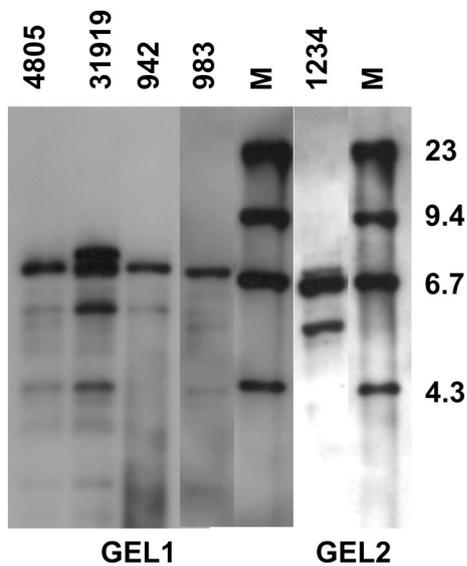


Fig. 2. Southern analysis of the *aaap* locus from several isolates of the *A. marginale* St. Maries strain. The image is a composite of two separate gels with λ HindIII markers shown in lanes 5 and 7 and DNA sizes in kb indicated on the right. Gel 1 (left) had two different exposure times. Empty lanes were removed to conserve space. Numbers on top indicate animal identification number.

dynamic variation with several subpopulations displaying different sized loci. The same blots were probed with *msp5* (Fig. 1B) to control for the amount and integrity of the DNA and complete digestion by the restriction endonucleases as evidenced by the single band seen with this hybridization. The dynamic nature of the subpopulations within the St. Maries strain was further investigated by Southern analysis using blood samples from several different calves that had been infected from the same source of stabilate. Fig. 2 shows a composite Southern of five different St. Maries strain infections. While the 5.6 kb and 6.7 kb locus sizes were present in several infections, the stoichiometry of these bands differed between the samples. Faint bands representing less prevalent subpopulations were also present with anywhere from 1 to 8 or more bands in a given infection. Together these data show the extensive variation of the *aaap* locus not only between strains, but also within a single strain.

3.2. Structural variation of the *aaap* locus is mediated by repeat sequences

Complete genome sequencing of the St. Maries strain provides sequence information for a single *aaap* locus. To analyze locus plasticity, four variants of the St. Maries strain and two variants of the Virginia strain locus were cloned without amplification when possible to avoid introducing errors due to the polymerase and sequenced. The sequences of three St. Maries strain clones derived from blood stage infection containing insert sizes of 4.5 kb, 4.6 kb, and 6.7 kb and one 6.7 kb clone from ISE6 cell culture were aligned with the previously reported 5.6 kb St. Maries *aaap* locus for comparison (Fig. 3A). Two Virginia strain clones of 8.1 kb and 6.3 kb (Fig. 3B) were also compared. The repetitive nature of this locus includes multiple repeat sequences both within and between genes. The 6.7 kb St. Maries *aaap* loci from blood and ISE6 cells were nearly identical and revealed a gene duplication event as compared with the 5.6 kb locus from the St. Maries genome. This duplication of the *alp2* gene was flanked by intergenic repeat regions consisting of a poly-G tract of 9–13 bp adjacent to an imperfect TCC repeat. This repeat sequence occurs between all genes in this locus as well as downstream of the *aaap* gene in the Florida and Virginia strains, but within the 3' end of the longer St. Maries *aaap* gene. In contrast to the expanded 6.7 kb St. Maries loci, the size difference in the 4.5 kb and 4.6 kb loci corresponded to deletion events which occurred

from the middle of *alp2* to the middle of *alp1*, resulting in gene fusions in both loci. Both fusions occurred in frame, were flanked by repeat sequences, and upstream and downstream sequences remained unaltered. The repeats bordering both fusions were located within the highly repetitive central regions of the genes.

Alignment of the 8.1 kb Virginia *aaap* locus and previously sequenced 7.7 kb Florida *aaap* locus (Dark et al., 2009) revealed similarities in both sequence and structure and each locus was composed of five genes: a duplicated *aaap* and an *alp1*, *alp2* and an *alp3* gene (Fig. 3B and C). The size difference between the 8.1 kb Virginia and 7.7 kb Florida loci corresponded primarily with insertions/deletions of intragenic repeat sequences. Alignments of the 8.1 kb and 6.3 kb *aaap* loci from the Virginia strain showed a 1840 bp deletion of *alp3* and the first *aaap* gene. This deletion was mediated by identical 116 bp sequences located at the 5' ends of both *alp3* and *alp2*.

3.3. Analysis of intragenic repeats

Comparisons of the deduced amino acid sequences in the Aaap protein family from the St. Maries (Brayton et al., 2005), Florida (Dark et al., 2009), Virginia, and Illinois (Stich et al., 2004) strains revealed conserved amino and carboxyl termini between similar proteins and a highly repetitive central region. The repetitive elements within the central region were classified into fourteen different “words,” eight of which contained the previously recognized ELKAIDAE motif (Stich et al., 2004), and were numbered for ease of visualization of the order of repeats. Despite low sequence identity within the Aaap protein family (Dark et al., 2009), common patterns were evident in the arrangement of repeats. For example, the Alp1 proteins in the St. Maries, Florida and Virginia strains contained nearly identical amino and carboxyl end sequences flanking the central polymorphic region. For example, the first 62 amino acids were 98% identical between these three sequences, while the last 185 amino acids were 85% identical. While the central repeat rich regions in the Alp1 proteins varied in length and sequence, they maintained a similar pattern of arranged words, such as the “2381238452” arrangement (Fig. 4). The St. Maries strain had the most unique sequences with some ELKAIDAE containing repeats unique to the Aaap and Alp2 proteins. Additionally, the carboxyl end of the St. Maries Aaap was unique and did not contain the characteristic terminal end sequence AAEVATPSTALGV that was present in all other proteins observed. The St. Maries *aaap* gene contained a poly-G tract which led to frameshift events, resulting in alternate termination sites for the Aaap protein product in the different versions of the locus that was sequenced from this strain.

3.4. *aaap* locus variation is reduced in tick infections

Locus plasticity in the St. Maries strain was examined in DNA isolated from infected *D. andersoni* ticks and ISE6 tick cell culture via Southern analysis (Fig. 5A). In contrast to the multiple bands observed in St. Maries genomic DNA isolated from red blood cells (Fig. 2), DNA isolated from infected ISE6 tick culture revealed a single band, showing a marked reduction in the subpopulation structure of this locus. The ISE6 *Ixodes scapularis* tick embryonic cell culture system is used to model tick infection, and is thought to be a surrogate for the midgut infectious stages (Noh et al., 2006). Southern analysis was also performed on DNA isolated from tick midgut and salivary gland tissues. The detection of the *aaap* locus in infected tick midgut and salivary gland tissues required DNA from 40 ticks to be pooled to obtain sufficient material for Southern analysis. The single band observed was the same size as that seen in the infected ISE6 cells; however, because the amount of *A. marginale* DNA within the tick samples was relatively low compared to the more abundant DNA from the large tick genome, it is possible that subpopulations are present below the level of detection. The single band indicated that most of the ticks were infected with *A. marginale* organisms containing the same locus structure. To establish

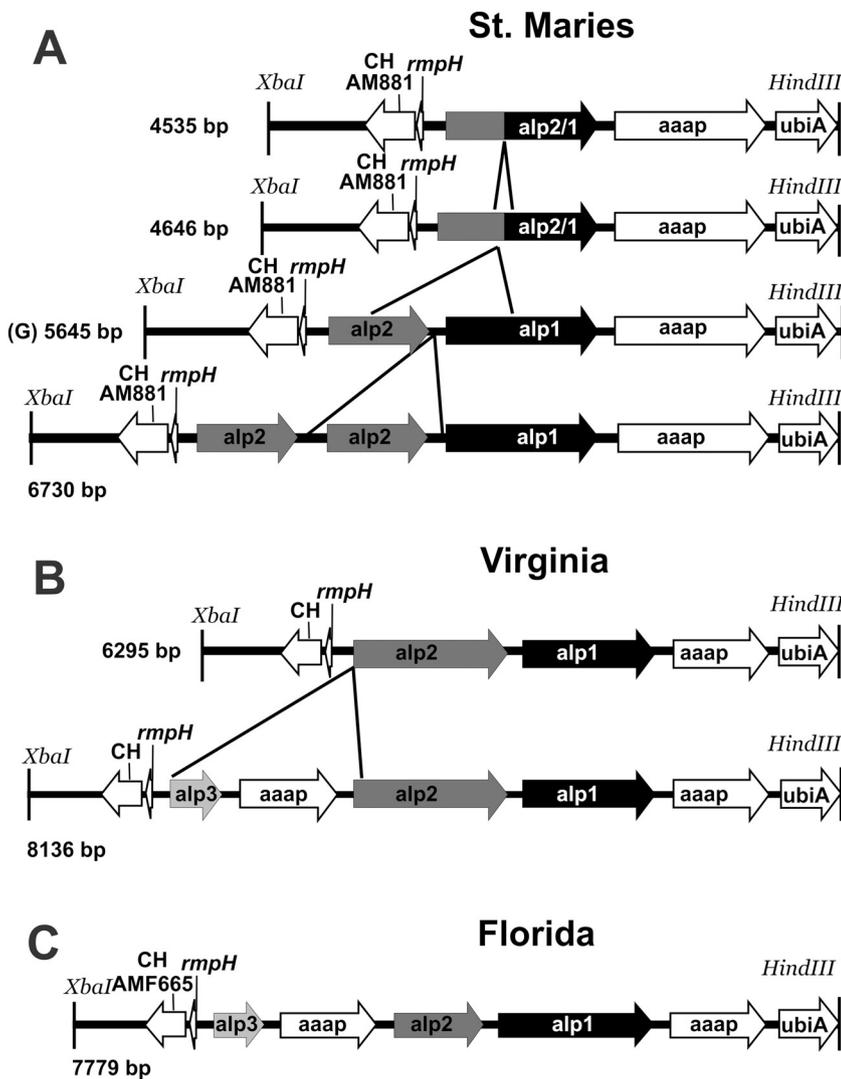


Fig. 3. Schematic representation of *aaap* locus variants. The horizontal bar represents the genome backbone for the locus spanning from conserved *XbaI* and *HindIII* sites and respective sizes on the left in bp. The orientation of each gene is indicated by the arrow head. Rearrangement events are shown by lines between the loci. Poly-G tract positions are denoted by CH. (A) Alignment of *aaap* loci from the St. Maries strain. Previously annotated *aaap* locus in GenBank (G) is shown aligned with three newly sequenced *aaap* loci. (B) Alignment of two sequenced *aaap* loci from the Virginia strain. (C) Schematic of previously annotated *aaap* locus from the Florida strain.

whether the restriction of locus structural variants was maintained after passage into a bovine host, locus structure was analyzed from calf no. 36361 which was infected from ISE6 cell culture. Southern analysis indicated that upon infection into the bovine host, subpopulations of *A. marginale* emerged containing *aaap* loci of different sizes. To confirm equal loading and integrity of *A. marginale* DNA, *msp5* was used as a control (Fig. 5B).

3.5. Transcriptional analysis of the *aaap* locus in infected erythrocytes and tick cell culture

Transcript data was obtained via RNA-Seq from St. Maries infected bovine erythrocytes and ISE6 tick cell culture and mapped to the St. Maries genome. In order to more accurately assess St. Maries transcript levels in ISE6 cells, the *aaap* locus cloned and sequenced from ISE6 cells was spliced in silico into the reference genome between the flanking *HindIII* and *XbaI* restriction sites. This was because the *aaap* locus annotated in the genome is not the dominant *aaap* locus (and may not even be present) in *A. marginale* infecting the tick cell. Similarly, of the dominant *aaap* loci present within St. Maries isolated from blood, the larger 6.7 kb *aaap* locus was spliced in silico into the reference genome for transcript mapping purposes. RNA-Seq reads obtained from *A. marginale* infected erythrocytes and ISE6 cells were then mapped to these modified reference genomes (Fig. 6). Reads mapped to the *aaap* locus from infected erythrocytes (Fig. 6A) and infected ISE6 cells

(Fig. 6B) showed that all four genes were expressed in both host environments. There are several gaps which indicate that transcription is not contiguous across the locus which may be due to insufficient coverage or a biological phenomenon. More gaps are seen in St. Maries transcriptome isolated from blood than from tick cells, particularly in *aaap* gene expression; however, it should be noted that there was twice the coverage obtained from tick cells than from blood.

To quantify gene expression levels, the reads per kilobase per million mapped reads (RPKM) were calculated for all genes within the *aaap* locus from data from both blood and tick stage samples as well as the mean RPKM value for all genes in the genome (Table 1). The RPKM values for St. Maries from blood stage infection show the average RPKM for all genes as 1008.6 while RPKM values for the *aaap* locus genes ranged from 92.0 to 203.0. Gene expression in the *aaap* locus from *A. marginale* infected tick cells was also much lower than the genome-wide average RPKM of 1165.0, with *alp1* having the highest RPKM at 198.3 while the *alp2* genes and *aaap* ranged from 41.7 to 86.0. While all the *aaap* locus genes were transcribed, transcript reads were 80% or more below the average RPKM values across the chromosome. In order to compare RPKM values between the blood and tick samples, the expression values were normalized. Almost all genes within the *aaap* locus were down regulated in *A. marginale* from tick culture with the exception of *alp1*, which was up-regulated. Overall, these data show that all genes within the *aaap* locus are expressed in both the bovine and tick hosts, but are expressed at a lower level than the average total gene

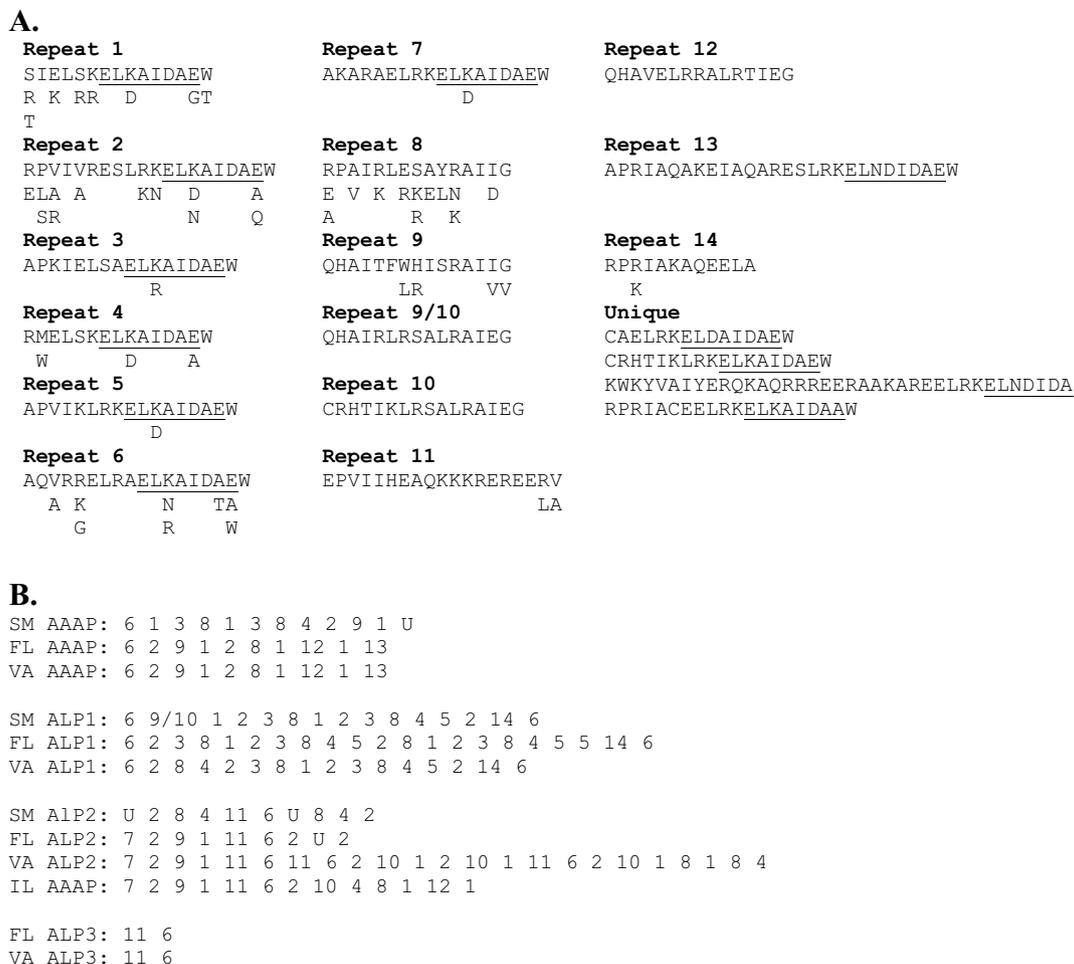


Fig. 4. Aaap repeat elements. (A) Sequence of repeat elements. Variations in repeat sequence are indicated by letters below primary repeat sequence. The ELKAIDAE motifs are underlined. (B) Order of repeat elements of the Aaap, Alp1, Alp2, and Alp3 proteins from the *A. marginale* St. Maries (SM), Florida (FL), Virginia (VA), and Illinois (IL) strains. Sequences unique to that particular protein are designated by a U. The Illinois strain Aaap protein was grouped with the Alp2 proteins as these repeat patterns were most similar.

expression in the chromosome. Differences in gene expression within this locus also occur depending on the host environment.

3.6. Aaap protein expression in infected erythrocytes and tick cell culture

Western blot analysis was performed to examine the expression of the Aaap proteins of the *A. marginale* St. Maries strain in both blood and tick samples. Proteins were isolated from blood samples collected during acute rickettsemia of St. Maries infected calf 1234 and from ISE6 cell culture infected with the St. Maries strain. To assess the expression profile of the Aaap proteins, protein extracts were separated by SDS-PAGE and immunoblotted with MAb Ana024D1 directed against Aaap (Fig. 7). To control for equal loading of organisms, MAb AnaF16C1 was used to detect Msp5, a constitutively expressed outer membrane protein that is highly conserved among *A. marginale* strains (Visser et al., 1992). Aaap proteins and Msp5 were detected in St. Maries infected bovine red blood cells and infected ISE6 cells, but not in uninfected samples. Three predominant Aaap protein bands were detected in the infected red blood cell lanes of the approximate molecular sizes 40, 42, and 46 kDa as well as fainter bands ranging in size from 30 to 46 kDa. In the infected ISE6 cell lane, only one protein band of 47 kDa was observed. The single protein expressed in St. Maries infected ISE6 cells was slightly larger in size than the predominantly expressed Aaap proteins within the bovine host.

3.7. Aaap and intragenic repeats in closely related species

The *aaap* gene was also found in *A. centrale* (GenBank: CP001759) and *A. ovis* (GenBank: CP015994). In *A. centrale*, there are two genes in

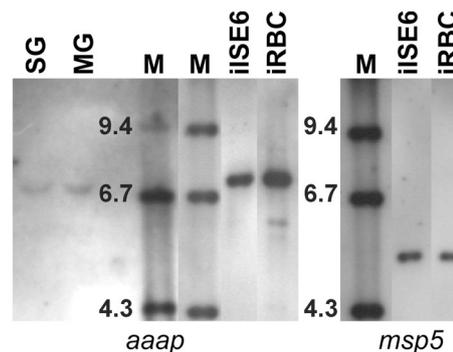


Fig. 5. Southern analysis of the *A. marginale* St. Maries *aaap* locus from tick and blood samples. Image is taken from the same gel with lanes moved horizontally only to conserve space (A) DNA was isolated from the infected salivary glands (SG) and midguts (MG) of *Dermacentor andersoni*, from infected ISE6 cells (iISE6), and infected bovine erythrocytes (iRBC) of animal 36361. (B) Comparable loading of DNA from iISE6 and infected erythrocytes was confirmed by probing for *msp5*. M indicates λ HindIII marker.

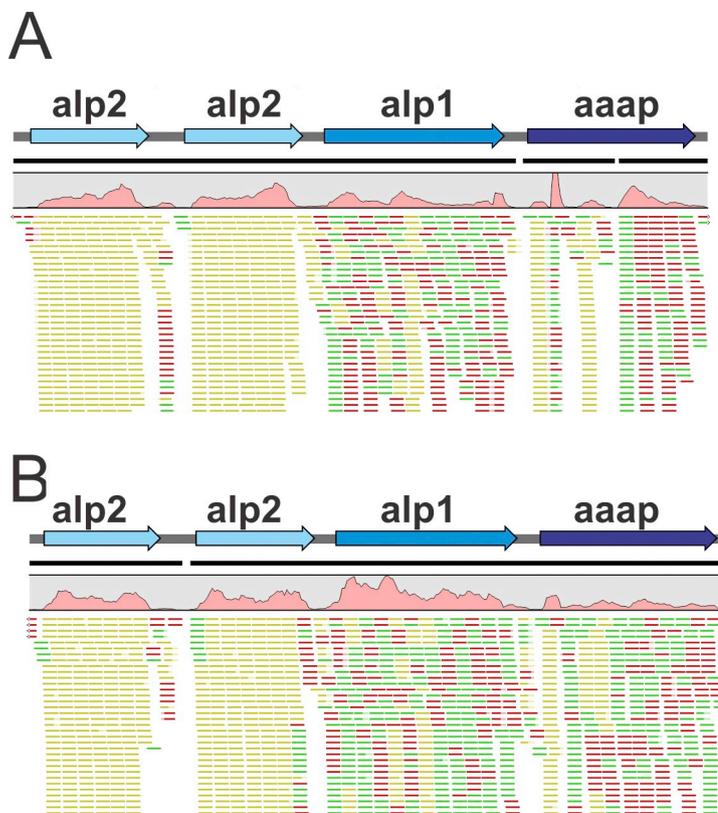


Fig. 6. Transcription profile of the *aaap* locus from infected bovine blood and ISE6 tick cell culture. A. Transcriptional mapping of *St. Maries* from infected blood to the longer (6.7 kb) sequenced *St. Maries aaap* locus. B. Transcriptional mapping of *St. Maries* from infected ISE6 tick culture cells to the *aaap* locus cloned and sequenced from infected tick cells. Unique reads are shown in red and green which map to the positive and negative sense strands. Reads shown in yellow indicate mapping to multiple locus positions such as reads mapping to the duplicated *alp2* genes. The black bar shows coverage across the locus with gaps as breaks in coverage. The pink histogram indicates the density of reads mapping to a particular position of the locus. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

the *A. marginale* syntenic locus that correspond to *alp1* and *alp2*, and there is no gene identified as *aaap* itself. There is a third Alp (*alp3*) in a distal location to the other two genes, however, this gene contains fewer ELKAIDA motifs. In *A. ovis*, there are also two genes in the syntenic locus, but these correspond to *aaap* and *alp1*. Again, there is a third gene found in a distal location, and this gene encodes a related motif “ELRAINA”.

3.8. Statistical analysis of the relationship between repeating motif patterns and Anaplasma

When examining the deduced amino acid sequences from all of the genes from the three species (*A. marginale*, *A. centrale*, and *A. ovis*), six similar repeating motifs become evident: ELDAIDA, ALKAIDA, ELKAIDA, ELRAIDA, ELRAIDE and ELRAINA, for simplicity referred to as the ELKAIDA motif. An exhaustive search was performed of the entire GenBank NR database, and no other bacteria were found to contain the ELKAIDA motif. The twelve genomes that were found all belong to

the genus *Anaplasma* with a higher sampling of *A. marginale* due to frequency of study of this pathogen. Motif repetition was again prevalent in all twelve organisms. As discussed earlier in the Methods section, based on the ratio of the number of single occurrences of the motif and the total number of amino acid combinations in an average protein sequence, the probability of a single motif of length 7 occurring randomly in a protein sequence is estimated to be vanishingly small, approximately 10^{-406} . This is for a single motif. In fact, the motifs occur multiple times in the *aaap* genes in *Anaplasma* and only in *Anaplasma*. This indicates that the appearance of the motifs are statistically significant in *Anaplasma*.

Table 1

A. marginale *St. Maries* transcript mapping in infected blood and tick cell culture.

	RPKM ^a	Total gene reads	Normalized RPKM
Bovine erythrocytes			
Chromosome	1008.58		1086.76
<i>alp2</i>	110.19	488	128.94
<i>alp2</i>	202.99	899	228.77
<i>alp1</i>	111.57	731	129.78
<i>aaap</i>	91.96	560	105.83
ISE6 tick culture			
Chromosome	1164.95		1086.76
<i>alp2</i>	44.04	375	40.49
<i>alp2</i>	85.95	744	78.19
<i>alp1</i>	198.28	2539	184.58
<i>aaap</i>	41.67	523	38.11

^a RPKM = reads per kilobase per million mapped reads.

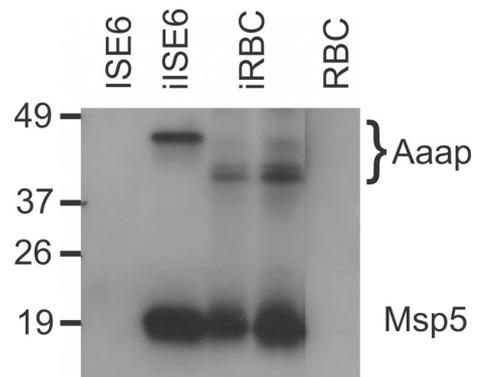


Fig. 7. Expression profile of Aaap proteins in *St. Maries* infected ISE6 cells (iISE6) and erythrocytes (iRBC). Two different amounts were loaded (1 × and 2 ×) in the infected erythrocyte lanes with Msp5 used as a positive control for comparable loading of the organisms between iRBC and iISE6 lanes. Noninfected ISE6 and erythrocytes (RBC) were used as negative controls. Protein marker indicating sizes is on the left.

4. Discussion

The hypothesis that repeat sequences mediate the expansion and contraction of the *aaap* locus is accepted. These repeats include both the polynucleotide tracts found between the *aaap* genes as well as intragenic sequences encoding the ELKAIDAE containing repeats. As repetitive sequences are believed to play a significant role in the generation of DNA duplications, they are often found flanking the 3' end, 5' end or both ends of the duplicated DNA segments (Meisel, 2009; Zhang, 2003). A common mechanism of DNA duplication occurs through unequal crossing over in which part of a gene, an entire gene, or several genes may be duplicated depending on the location of the cross over (Zhang, 2003). While the poly-G intergenic repeats in the *aaap* locus appear to facilitate rearrangement events of whole genes, they may also function in gene regulation. Homo and heteropolymeric nucleotide repeat tracts are common mutational hot spots as they are specifically vulnerable to simple slip-strand mispairing, leading to a reversible expansion or contraction of the repeating unit (van der Woude and Baumler, 2004; Moxon et al., 2006; van Belkum et al., 1998; Willems et al., 1990; van der Ende et al., 1995; van Ham et al., 1993). Scanning of available whole genome sequences shows there are approximately 20 poly-G tracts of 9 or more bases in the St. Maries and Florida strain genomes. These tracts displayed a bias for the leading strand, and were often found to be closely associated with the 5' end of genes. Three of the 21 St. Maries poly-G tracts are found within the *aaap* locus.

Sequence alignments of the *aaap* locus from the St. Maries, Florida, and Virginia strains showed deletions and insertions within genes that were mediated by the intragenic repeats inherent to the *aaap* gene family. Gene expression and functionality are influenced by intragenic DNA repeats where expansion or contraction of the repeating unit can directly affect the gene product and possibly result in phenotypic changes (Li et al., 2004; Kashi and King, 2006; Karlin et al., 2002; Verstrepen et al., 2005; Fondon 3rd and Garner, 2004). While rearrangements occurred between these intragenic repeats, they did not appear to be random as patterns in the sequence of repeats were apparent (Fig. 4). Intragenic repeat-mediated rearrangements were also evidenced by large deletion events. Two separate deletion events observed in the St. Maries *aaap* locus each resulted in a gene fusion of *alp2* and *alp1* which would lead to an altered protein product, if expressed (Fig. 3A). While this kind of rearrangement would result in a new gene product, it would likely retain the transcriptional regulation of the original *alp2* as the upstream regions containing the promoter and regulatory sequences remained unchanged. Similarities between the more conserved 5' and 3' ends as well as the common repeat segments found in this gene family suggest they may be paralogs that diverged from initial duplication events (Zhang, 2003). The retention of the deduced amino acid ELKAIDAE motif among all repeats indicated a strong selection for a common function.

The dynamic nature of the *aaap* locus may provide a source of adaptability not only in the bovine host, but also in the tick vector. While morphological studies have shown the presence of similar appendage structures attached to the midgut lumen of *D. andersoni* nymphs, it remains unknown as to whether the appendages are involved in tick infection or transmission (Kocan et al., 1984). Transcriptional analysis of the St. Maries *aaap* loci showed expression of all genes within this locus in both bovine and tick hosts, albeit significantly lower than the overall average gene expression of the chromosome. Transcript data of the genes in the *aaap* locus showed similar levels of expression within the bovine host with *alp2* being the most highly expressed. In contrast, gene expression in the *aaap* locus from infected tick cells was reduced with the exception of *alp1*. Breaks in coverage indicated that these genes are unlikely to be polycistronically transcribed. Western blot analysis of ISE6 tick cell culture and bovine erythrocytes infected with the St. Maries strain showed the expression of Aaap proteins (Fig. 6). The expression profile of the Aaap proteins differed in that at least two proteins were expressed in the bovine host whereas

only a single protein was detected in the tick cells. This single protein of approximately 47 kDa, consistent with the predicted size of Alp1, is larger than the predominantly expressed proteins in the red blood cells, indicating a regulatory shift in protein expression from the bovine host to the tick cell. A study using a proteomic approach to identify proteins expressed by the St. Maries strain in ISE6 cells identified the presence of an Aaap-like protein (Ramabu et al., 2010). Identification of a specific *aaap* locus protein is difficult due to not only the repetitive nature in this protein family, but also the propensity for rearrangement events leading to variable protein products.

As the protein expression profile varied dramatically between infected tick culture cells and infected bovine red blood cells, Southern analysis revealed that the subpopulations prevalent in erythrocytes seemed to have disappeared within the tick host, leaving a single *aaap* locus structure. The single *aaap* locus seen in the ISE6 cell line and midgut and salivary glands indicates a reduction of population structure of the *aaap* locus while in the tick vector, suggesting a reduction in variants may be due to a specific *aaap* locus having the greatest fitness within the tick vector. This simplification of the *aaap* population within the tick could also account for fewer proteins being expressed. As only a single protein variant was observed, this variant may be host specific and possibly involved directly or indirectly by interaction with unknown factors in infectivity, survival or transmission. As other variants were not expressed, this may imply they are less beneficial and thus were selected against. Interestingly, locus subpopulations began to reappear with the inoculation of the St. Maries strain from the ISE6 cells into a naïve calf. The reemergence of locus subpopulations within the bovine host from the tick vector suggests that multiple protein variants are more advantageous. As the specific identification of the proteins expressed in erythrocytes remains unknown, the possibility that only a single protein is being expressed from each cell but the population contains many variants cannot be excluded.

While little is currently known as to the role and impact the formation of the inclusion appendages has on the parasite's life cycle, our research has shown the *aaap* locus to be a very dynamic region in the *A. marginale* genome which has been retained and diversified in the face of reductive evolution. The plasticity observed in this locus appeared to be mediated by the numerous repeat elements inherent to this locus, including the ELKAIDAE containing repeats. These proteins have been shown to be differentially expressed between the bovine and tick hosts, suggesting an altered expression profile in response to a changing environment. This host-specific response was also noted in a reduced number of *aaap* locus variants when passed to the arthropod vector from the bovine host. Taken together, these findings warrant further investigations into what role the *aaap* locus and its encoded proteins may play in the infection and persistence of disease for *A. marginale*.

Acknowledgements

We thank Ralph Horn, James Allison and Beverly Hunter for excellent technical assistance.

This work was supported by National Institutes of Health (Grant # AI44005) and the U.S. Department of Agriculture CSREES (Grant #2005-35604-15440). The funders had no role in the study design or decision to submit the manuscript to this journal.

References

- Boynton, W., 1932. Further observations on anaplasmosis. *Cornell Vet.* 22, 10–28.
- Brayton, K.A., Kappmeyer, L.S., Herndon, D.R., Dark, M.J., Tibbals, D.L., Palmer, G.H., McGuire, T.C., Knowles Jr., D.P., 2005. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc. Natl. Acad. Sci. U. S. A.* 102, 844–849.
- Dark, M.J., Herndon, D.R., Kappmeyer, L.S., Gonzales, M.P., Nordeen, E., Palmer, G.H., Knowles Jr., D.P., Brayton, K.A., 2009. Conservation in the face of diversity: multi-strain analysis of an intracellular bacterium. *BMC Genomics* 10, 16. <https://doi.org/10.1186/1471-2164-10-16>.
- Eriks, I.S., Stiller, D., Goff, W.L., Pantan, M., Parish, S.M., McElwain, T.F., Palmer, G.H.,

1994. Molecular and biological characterization of a newly isolated *Anaplasma marginale* strain. *J. Vet. Diagn. Investig.* 6, 435–441.
- Espana, E., Espana, C., 1963. *Anaplasma marginale*. II. Further studies of morphologic features with phase contrast and light microscopy. *Am. J. Vet. Res.* 24, 713–722.
- Felsheim, R.F., Chavez, A.S., Palmer, G.H., Crosby, L., Barbet, A.F., Kurtti, T.J., Munderloh, U.G., 2010. Transformation of *Anaplasma marginale*. *Vet. Parasitol.* 167, 167–174.
- Fondon 3rd, J.W., Garner, H.R., 2004. Molecular origins of rapid and continuous morphological evolution. *Proc. Natl. Acad. Sci. U. S. A.* 101, 18058–18063.
- Franklin, T.E., Redmond, H.E., 1958. Observations on the morphology of *Anaplasma marginale* with reference to projections or tails. *Am. J. Vet. Res.* 19, 252–253.
- Hidalgo, R.J., Palmer, G.H., Jones, E.W., Brown, J.E., Ainsworth, A.J., 1989. Infectivity and antigenicity of *Anaplasma marginale* from tick cell culture. *Am. J. Vet. Res.* 50, 2033–2036.
- Karlin, S., Brocchieri, L., Bergman, A., Mrazek, J., Gentles, A.J., 2002. Amino acid runs in eukaryotic proteomes and disease associations. *Proc. Natl. Acad. Sci. U. S. A.* 99, 333–338.
- Kashi, Y., King, D.G., 2006. Simple sequence repeats as advantageous mutators in evolution. *Trends Genet.* 22, 253–259.
- Kocan, K.M., Venable, J.H., Brock, W.E., 1978a. Ultrastructure of anaplasma inclusions (Pawhuska isolate) and their appendages in intact and hemolyzed erythrocytes and in complement-fixation antigen. *Am. J. Vet. Res.* 39, 1123–1130.
- Kocan, K.M., Venable, J.H., Hsu, K.C., Brock, W.E., 1978b. Ultrastructural localization of anaplasma antigens (Pawhuska isolate) with ferritin-conjugated antibody. *Am. J. Vet. Res.* 39, 1131–1135.
- Kocan, K.M., Ewing, S.A., Hair, J.A., Barron, S.J., 1984. Demonstration of the inclusion appendage of *Anaplasma marginale* in nymphal *Dermacentor andersoni*. *Am. J. Vet. Res.* 45, 1800–1807.
- Kocan, K.M., de la Fuente, J., Guglielmo, A.A., Melendez, R.D., 2003. Antigens and alternatives for control of *Anaplasma marginale* infection in cattle. *Clin. Microbiol. Rev.* 16, 698–712.
- Kreier, J.P., Ristic, M., 1963. Anaplasmosis. X. Morphologic characteristics of the parasites present in the blood of calves infected with the Oregon strain of *Anaplasma marginale*. *Am. J. Vet. Res.* 24, 676–687.
- Kuttler, K.L., Winward, L.D., 1984. Serologic comparisons of 4 *Anaplasma* isolates as measured by the complement-fixation test. *Vet. Microbiol.* 9, 181–186.
- Li, Y.C., Korol, A.B., Fahima, T., Nevo, E., 2004. Microsatellites within genes: structure, function, and evolution. *Mol. Biol. Evol.* 21, 991–1007.
- McGuire, T.C., Palmer, G.H., Goff, W.L., Johnson, M.I., Davis, W.C., 1984. Common and isolate-restricted antigens of *Anaplasma marginale* detected with monoclonal antibodies. *Infect. Immun.* 45, 697–700.
- Meisel, R.P., 2009. Repeat mediated gene duplication in the *Drosophila pseudoobscura* genome. *Gene* 438, 1–7.
- Miller, D.F., Yan, P.S., Buechlein, A., Rodriguez, B.A., Yilmaz, A.S., Goel, S., Lin, H., Collins-Burrow, B., Rhodes, L.V., Braun, C., Pradeep, S., Rupaimoole, R., Dalkilic, M., Sood, A.K., Burrow, M.E., Tang, H., Huang, T.H., Liu, Y., Rusch, D.B., Nephew, K.P., 2013. A new method for stranded whole transcriptome RNA-seq. *Methods* 63, 126–134.
- Mortazavi, A., Williams, B.A., McCue, K., Schaeffer, L., Wold, B., 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat. Methods* 5, 621–628.
- Moxon, R., Bayliss, C., Hood, D., 2006. Bacterial contingency loci: the role of simple sequence DNA repeats in bacterial adaptation. *Annu. Rev. Genet.* 40, 307–333.
- Munderloh, U.G., Liu, Y., Wang, M., Chen, C., Kurtti, T.J., 1994. Establishment, maintenance and description of cell lines from the tick *Ixodes scapularis*. *J. Parasitol.* 80, 533–543.
- Munderloh, U.G., Blouin, E.F., Kocan, K.M., Ge, N.L., Edwards, W.L., Kurtti, T.J., 1996. Establishment of the tick (*Acari: Ixodidae*)-borne cattle pathogen *Anaplasma marginale* (*Rickettsiales: Anaplasmataceae*) in tick cell culture. *J. Med. Entomol.* 33, 656–664.
- Noh, S.M., Brayton, K.A., Knowles, D.P., Agnes, J.T., Dark, M.J., Brown, W.C., Baszler, T.V., Palmer, G.H., 2006. Differential expression and sequence conservation of the *Anaplasma marginale msp2* gene superfamily outer membrane proteins. *Infect. Immun.* 74, 3471–3479.
- Palmer, G.H., Knowles Jr., D.P., Rodriguez, J.L., Gnad, D.P., Hollis, L.C., Marston, T., Brayton, K.A., 2004. Stochastic transmission of multiple genotypically distinct *Anaplasma marginale* strains in a herd with high prevalence of *Anaplasma* infection. *J. Clin. Microbiol.* 42, 5381–5384.
- Pierle, S.A., Dark, M.J., Dahmen, D., Palmer, G.H., Brayton, K.A., 2012. Comparative genomics and transcriptomics of trait-gene association. *BMC Genomics* 13, 669.
- Ramabu, S.S., Ueti, M.W., Brayton, K.A., Baszler, T.V., Palmer, G.H., 2010. Identification of *Anaplasma marginale* proteins specifically upregulated during colonization of the tick vector. *Infect. Immun.* 78, 3047–3052.
- Ristic, M., Carson, C.A., 1977. Methods of immunoprophylaxis against bovine anaplasmosis with emphasis on use of the attenuated *Anaplasma marginale* vaccine. *Adv. Exp. Med. Biol.* 93, 151–188.
- Scoles, G.A., Ueti, M.W., Noh, S.M., Knowles, D.P., Palmer, G.H., 2007. Conservation of transmission phenotype of *Anaplasma marginale* (*Rickettsiales: Anaplasmataceae*) strains among *Dermacentor* and *Rhipicephalus* ticks (*Acari: Ixodidae*). *J. Med. Entomol.* 44, 484–491.
- Stich, R., Kocan, K.M., Damian, R.T., Fechheimer, M., 1997. Inclusion appendages associated with the intraerythrocytic rickettsial parasite *Anaplasma marginale* are composed of bundled actin filaments. *Protoplasma* 199, 93–98.
- Stich, R.W., Olah, G.A., Brayton, K.A., Brown, W.C., Fechheimer, M., Green-Church, K., Jittapalpong, S., Kocan, K.M., McGuire, T.C., Rurangirwa, F.R., Palmer, G.H., 2004. Identification of a novel *Anaplasma marginale* appendage-associated protein that localizes with actin filaments during intraerythrocytic infection. *Infect. Immun.* 72, 7257–7264.
- van Belkum, A., Scherer, S., van Alphen, L., Verbrugh, H., 1998. Short-sequence DNA repeats in prokaryotic genomes. *Microbiol. Mol. Biol. Rev.* 62, 275–293.
- van der Ende, A., Hopman, C.T., Zaat, S., Essink, B.B., Berkhout, B., Dankert, J., 1995. Variable expression of class 1 outer membrane protein in *Neisseria meningitidis* is caused by variation in the spacing between the –10 and –35 regions of the promoter. *J. Bacteriol.* 177, 2475–2480.
- van der Woude, M.W., Baumler, A.J., 2004. Phase and antigenic variation in bacteria. *Clin. Microbiol. Rev.* 17, 581–611 (table of contents).
- van Ham, S.M., van Alphen, L., Mooi, F.R., van Putten, J.P., 1993. Phase variation of *H. influenzae* fimbriae: transcriptional control of two divergent genes through a variable combined promoter region. *Cell* 73, 1187–1196.
- Verstrepen, K.J., Jansen, A., Lewitter, F., Fink, G.R., 2005. Intragenic tandem repeats generate functional variability. *Nat. Genet.* 37, 986–990.
- Visser, E.S., McGuire, T.C., Palmer, G.H., Davis, W.C., Shkap, V., Pipano, E., Knowles Jr., D.P., 1992. The *Anaplasma marginale msp5* gene encodes a 19-kilodalton protein conserved in all recognized *Anaplasma* species. *Infect. Immun.* 60, 5139–5144.
- Willems, R., Paul, A., van der Heide, H.G., ter Avest, A.R., Mooi, F.R., 1990. Fimbrial phase variation in *Bordetella pertussis*: a novel mechanism for transcriptional regulation. *EMBO J.* 9, 2803–2809.
- Zhang, J., 2003. Evolution by gene duplication: an update. *Trends Ecol. Evol.* 18, 292–298.