COMPARATIVE GENOMICS OF *ANAPLASMA MARGINALE* – A PRELIMINARY EXAMINATON OF FACTORS INVOLVED IN TICK TRANSMISSION

By

MICHAEL JAMES DARK

A dissertation submitted in partial fulfillment of
the requirements for the degree of

DOCTOR OF PHILOSOPHY

WASHINGTON STATE UNIVERSITY
College of Veterinary Medicine

December 2008

To the faculty of Washington State University:

The members of the committee appointed to examine the dissertation of MICHAEL JAMES DARK find it satisfactory and recommend that it be accepted.

____
Chair

____

____

____

ACKNOWLEDGEMENTS

This work would not have been possible without the help and support of my major advisor, Kelly Brayton. I could not have asked for a better mentor.

Thanks also go to Guy Palmer, Don Knowles Jr., and Fred Rurangirwa, whose contributions as mentors and committee members were crucial to this work. This work was facilitated by the excellent technical support was provided by Beverly Hunter, Xiaoya Cheng, and Ralph Horn.

I also gratefully acknowledge my colleagues, including James Stanton, Joshua Daniels, Danielle Nelson, and Seth Harris, whose friendship and support allowed me to finish this work.

Finally, I acknowledge the financial support of the National Institutes of Health through my Mentored Clinical Scientist Research Award (K08-AI64162).

COMPARATIVE GENOMICS OF *ANAPLASMA MARGINALE* – A PRELIMINARY EXAMINATON OF FACTORS INVOLVED IN TICK TRANSMISSION

Abstract

by Michael James Dark, D.V.M., Ph D.
Washington State University
December 2008

Chair: Kelly A. Brayton

Anaplasma marginale is an obligate intraerythrocytic parasite of cattle, causing cyclic anemia, decreased milk and meat production, and occasionally death. Epidemiological persistence requires mechanical or biological transmission. Biological transmission via ixodid ticks is the most significant means of transmission, with *Dermacentor andersoni* being the most common agent in the Pacific Northwest. Previous studies have found several strains that are not transmissible by *D. andersoni*; however, research to date has not linked this phenotype with a molecular cause. Because phenotype is encoded by genotype, we have undertaken a comparative genomic approach to identify genes involved in tick transmission.

We have completed genome sequencing of the Florida strain, which is not transmissible by *D. andersoni*, and compared it to the previously sequenced St. Maries strain genome, to associate genetic differences with the observed phenotypic differences. DNA from three other strains (two tick-transmissible and one non-

transmissible) were subjected to pyrosequencing and assembled, to increase our power of association. These studies show *A. marginale* has a closed-core genome with a high degree of polymorphisms between strains.

While global comparisons are ongoing, we also tested the only *A. marginale* gene currently associated with tick transmission, to determine its role in determining this phenotype. Specifically, the repeat region of msp1a was implicated in adhesion to tick cells. Two strains were identified with the same msp1α genotype, but different transmission phenotypes. We tested the South Idaho and Mississippi strains, both with the msp1α genotype DDDDE, for transmission status with two tick species, *D. andersoni* and *Rhipicephalus microplus*. The Mississippi strain was not transmitted by either tick species, while the South Idaho strain was transmitted with *D. andersoni* but not *R. microplus*. South Idaho infection levels were significantly higher in *D. andersoni* ticks than in any other strain/tick combinations, implicating infection level as a factor in transmissibility. Our data suggests that the msp1a repeats do not play a role in the differential transmissibility of the South Idaho and Mississippi strains.

Ongoing comparative analysis will produce a list of candidate genes involved in tick transmission for testing and development of transmission-blocking vaccines.

TABLE OF CONTENTS

Page

ACKNOWLEDGEMENTS ___ **III**

ABSTRACT ___ **IV**

LIST OF TABLES ___ **VIII**

LIST OF FIGURES ___ **IX**

DEDICATION ___ **X**

CHAPTER 1 ___ **1**

- ABSTRACT ___ 1
- BACKGROUND ___ 2
- RESULTS ___ 5
- DISCUSSION ___ 9
- CONCLUSIONS ___ 13
- METHODS ___ 14
- REFERENCES ___ 18
- TABLES ___ 26
- FIGURES ___ 32

CHAPTER 2 **37**
ABSTRACT 37
INTRODUCTION 38
MATERIALS AND METHODS 40
RESULTS 43
DISCUSSION 45
REFERENCES 48
TABLES 52
FIGURES 53

LIST OF TABLES

CHAPTER 1

TABLE 1 - COMPARISON OF THE ST. MARIES AND FLORIDA GENOME FEATURES ____ 26

TABLE 2 - IDENTITY BETWEEN DEDUCED AAAP AMINO ACID SEQUENCES FROM THE ST. MARIES AND FLORIDA STRAINS ____ 26

TABLE 3 - PYROSEQUENCING RESULTS FOR THREE STRAINS OF *ANAPLASMA MARGINALE* ____ 26

TABLE 4 – PYROSEQUENCING GAPS ____ 27

TABLE 5 – SPECIES COMPARED IN SNP ANALYSIS ____ 29

CHAPTER 2

TABLE 1. ANALYSIS OF SALIVARY GLAND AND MIDGUT INFECTION STATUS VIA MSP5 SOUTHERN BLOTTING AFTER AQUISITION FEEDING ____ 52

TABLE 2. TRANSMISSION FEEDING RESULTS. ____ 52

LIST OF FIGURES

CHAPTER 1

FIGURE 1 - COMPARISON OF THE LEVEL OF SNP DIVERSITY AMONG SEQUENCED GENOMES. 32

FIGURE 2 - PHYSICAL MAP OF THE MSP2 FUNCTIONAL PSEUDOGENES IN THE ST. MARIES (STM) AND FLORIDA (FL) STRAINS. 33

FIGURE 3 - SOUTHERN BLOT OF THE AAAP LOCUS IN *A. MARGINALE* STRAINS. 33

FIGURE 4 - DISTRIBUTION OF GAPS IN THE THREE PYROSEQUENCED GENOMES. 34

FIGURE 5 - DISTRIBUTION OF SNPS IN FOUR STRAINS COMPARED TO THE FLORIDA STRAIN. 35

FIGURE 6 – SNP DISTRIBUTION IN THREE SPECIES 36

CHAPTER 2

FIGURE 1. *A. MARGINALE* INFECTION LEVELS 53

DEDICATION

This dissertation is dedicated to my wife, Carolyn. Her dedication and persistence are a continuing inspiration to me.

Thank you for sticking it out through four years apart and two years together.

It is also dedicated to my family, for their love and support through this long process.

CHAPTER 1

CONSERVATION IN THE FACE OF DIVERSITY: MULTISTRAIN ANALYSIS OF AN INTRACELLULAR BACTERIUM

ABSTRACT

With the recent completion of numerous sequenced bacterial genomes, notable advances have been made in understanding the level of conservation between various species. However, relatively little is known about the genomic diversity among strains. We determined the complete genome sequence of the Florida strain of *Anaplasma marginale*, and near complete (>96%) sequences for an additional three strains, for comparative analysis with the previously fully sequenced St. Maries strain genome. These comparisons revealed that *A. marginale* has a closed-core genome with few highly plastic regions, which include the *msp2* and *msp3* genes, as well as the *aaap* locus. Comparison of the Florida and St. Maries genome sequences found that SNPs comprise 0.8% of the longer Florida genome, with 33.5% of the total SNPs between all five strains present in at least two strains and 3.0% of SNPs present in all strains except Florida. Comparison of genomes from three strains of *Mycobacterium tuberculosis, Bacillus anthracis,* and *Nessieria meningiditis,* as well as four *Chlamydophila pneumoniae* strains found that 98.8%-100% of SNPs are unique to each strain, suggesting *A. marginale*, with 76.0%, has an intermediate level of strain-specific SNPs. Comparison of genomes from other organisms revealed variation in diversity that did not segregate with the environmental niche the bacterium occupies, ranging from 0.00% to 8.00% of the larger pairwise-compared genome. Analysis of multiple *A. marginale*

strains suggests intracellular bacteria have more variable SNP retention rates than previously reported, and may have closed-core genomes in response to the host organism environment and/or reductive evolution.

BACKGROUND

While the recent boom in genome sequencing projects has provided a wealth of information about bacterial metabolism and evolution, we know little about interstrain variation. A firm understanding of the rates and sites of variation is useful in determining genotypic differences associated with phenotypic traits and in formulating control strategies for a number of pathogens. Further, knowledge about the pan-genome of organisms will aid in determining the core genomic requirements, as well as shed more light on events that occur in the various environmental niches bacteria occupy.

Most studies of bacterial diversity to date have either utilized specific genomic loci [1, 2] or have examined metagenomics of specific environmental niches [3, 4]. While these types of studies help elucidate the extent of diversity, there is still a key component that has not yet been investigated – a measurement of diversity within bacterial species. Obtaining a true measure of species diversity is difficult, as the strains selected for whole-genome sequencing are generally chosen to examine a particular phenotypic trait, subjecting any resultant measures of diversity to selection bias.

The level of interstrain diversity can have a significant impact on the direction of research. Selection of pathogen strains for sequencing is typically based on differences

in virulence [5], host preference [6], or tissue tropism [7]. Using these selection criteria may artificially skew the level of diversity in the studied genome sequences, resulting in a biased level of diversity which does not accurately reflect the true genetic diversity of the species. However, since the diversity among strains has only been examined in a small number of species, determining if there is a skew is difficult. For example, analysis of several genome sequences of *Bacillus anthracis* found a low number of single-nucleotide polymorphisms (SNPs) [8], which led to development of other techniques for examining the epidemiology of outbreak strains [9]. *B. anthracis* is an example of a “closed core” genome – that is, after sequencing several strains (four for *B. anthracis*), no strain-specific genes are added to the pan-genome [10], which may be a result of a clonal population split of *B. anthracis* from *B. cereus*. Thus, the closed-core genome may be the result of small evolutionary distance, and may be a rare finding for organisms with larger evolutionary distance. The alternative is an “open core” genome, where each new sequenced strain adds at least one unique gene to the pan-genome. This is exemplified by *Streptococcus agalactiae*, which has approximately 30 new strain-specific genes for each additional genome sequenced, regardless of the total number of strains compared.

What influences the pan-genome? Is the pan-genome content fixed, or does it drift with time? Do all non-clonal populations have open-core genomes, or is this influenced by the environment a bacterium occupies? While answering all of these questions will require sequencing many more genomes, *Anaplasma marginale* makes an excellent system for studying the last question for a number of reasons. *A. marginale*

is a member of the order *Rickettsiales* and a well-established obligate intracellular bacterial model. *A. marginale* is the most globally prevalent vector-borne pathogen of cattle, causing cyclic anemia, decreased production, and possibly death [11]. A previous genome sequence for the St. Maries strain [11] establishes that this organism has a small genome size due to reductive evolution, and is related to several other intracellular pathogens, including those in the genera *Anaplasma*, *Ehrlichia*, and *Rickettsia* [12]. In addition, *A. marginale* has a number of characterized strains, with each strain defined by *msp1α* genotype [13, 14]. While previous studies have utilized specific genes to examine differences between these strains [15-17], no studies have examined the species diversity of *A. marginale*. A number of studies have described strains that vary in geographic location and phenotypic traits [15, 18, 19], and these are available for determination of the true level of genetic diversity in this species, subsequent analysis of the status of core genes, and determination if these are correlated with the intracellular lifestyle, geographic location, tick-transmissible status, or other characteristics of these organisms.

To answer this question, we obtained genome sequences for four strains of *A. marginale* that have differing abilities to be transmitted by *Dermacentor andersoni*, with each phenotype represented by at least two geographically distinct isolations. We sequenced the Florida strain to completion using a BAC-based clone by clone approach, and obtained high coverage genome sequence data for three additional strains. The resulting DNA sequences were then analyzed and compared to both the previously sequenced St. Maries genome [11], as well as other bacterial species in the

Order *Rickettsiales*. Further, the genomes of several other non-rickettsial bacteria were examined with similar genome comparison techniques to determine if diversity and pan-genome content are related to pathogenicity or an intracellular lifestyle.

RESULTS

Microbial genome diversity

Previous studies [6, 20] have shown high levels of variation between the genomes of different strains of obligate intracellular bacteria. To compare the levels found in *A. marginale* to other genomes, similar comparisons were made for organisms meeting the following criteria: organisms with 1) a single chromosome, 2) more than one sequenced strain, and 3) assembled and finished genome sequences deposited in Genbank, including free-living, facultative intracellular, and obligate intracellular bacteria (Figure 1). Single-factor analysis of variance (ANOVA) finds no significant differences in the level of variation between obligate intracellular, facultative, and free-living bacteria. The number of SNPs ranged from 0.00% to 8.00% of the larger genome, with significant intraspecies and intragenera variation.

General genome features and comparison of the St. Maries and Florida strains

The *Anaplasma marginale* Florida strain genome is composed of a single 1,202,435 bp circular chromosome predicted to contain 942 coding sequences (CDS) (Table 1). Similar to most other previously sequenced *Anaplasmataceae*, there are no plasmids and no identifiable insertion sequences. Compared to the previously sequenced St. Maries strain genome [11], there are seven fewer CDSs despite the larger genome size, due primarily to differences in split open reading frames (ORFs)

and annotation differences. The high degree of synteny between these two strains is disrupted by two inversions; one approximately 30kb long is flanked by repeat elements (*msp3* pseudogenes), while the other is a single gene flanked by short duplicated hypothetical genes.

Split ORFs were first described in the *Rickettsia conori* genome [21], and are postulated to represent genes that are in the first stage of reductive evolution. The idea that these ORFs have split recently is consistent with the findings in *Anaplasma*, as different ORFs are split in the two completely sequenced strains. The four split ORFs annotated in the St. Maries genome (*mutL*, *murC*, *aatA*, and *aspS*) [11] are intact in the Florida genome, and two tandem genes annotated as hypothetical in the St. Maries genome (AM574 and AM576) are fused in the Florida genome (AMF_437). Only one split ORF, *petA*, is found in the Florida genome. Four small ORFs in the St. Maries genome (AM380, AM395, AM974, and AM976), ranging in size from 204bp to 378bp are not present in the Florida genome. These ORFs are flanked by repetitive DNA sequences, and appear to be missing due to recombination events.

Genes mediating genome plasticity

The *msp2* superfamily is a group of related *A. marginale* genes encoding surface proteins [11]. *Msp2* encodes a highly antigenic protein that varies over time during infection by gene conversion of functional pseudogenes into a single expression site, to create new antigenic variants capable of evading the existing immune response. Compared to the St. Maries genome, the Florida genome has one additional *msp2*

functional pseudogene. Of the eight Florida *msp2* functional pseudogenes, four are identical to those in the St. Maries genome. The Florida genome has two sets of duplicated functional pseudogenes, TTV 4F15 / TTV 1O6 and KAV 4F15 / KAV 1F20 (Figure 2); while St. Maries was found to have duplicated functional pseudogenes, this was not noted in a functional pseudogene-targeted examination of other strains [22]. Florida has a set of duplicated functional pseudogenes in the same genome positions as St. Maries (2 / 3H1 in St. Maries, and KAV 4F15 / KAV 1F20 in Florida). As obligate intracellular bacteria are not thought to undergo lateral gene transfer, identical functional pseudogenes indicates the sequence is either evolutionarily conserved or has been selected independently in both strains due to a fitness advantage. Interestingly, both copies in Florida have a change encoding 15 amino acids at the 5' end of the hypervariable region compared to their St. Maries counterparts; either both strains duplicated a functional pseudogene after this change occurred in an ancestral strain, or both copies in one of the strains acquired identical changes after the ancestral strain duplicated the original functional pseudogene. In contrast, only two of the seven MSP3 functional pseudogenes are identical between Florida and St. Maries (*msp3* C / *msp3*-1, and *msp3* 4L1 / *msp3* 6). The *omp1-15* genes are present in both genomes, with a high degree of conservation between the predicted amino acid sequences (85.3-100% identity) as previously reported [17].

Aaap **gene family**

The *aaap* gene was first recognized and characterized as an *Anaplasma* appendage associated protein [23]. Subsequently, additional related genes were

identified that appear to be tandemly-duplicated copies that have diverged to have relatively low levels of sequence identity (Table 2). There is expansion of this locus in the Florida strain relative to the St. Maries strain, with a duplicated copy of the *aaap* gene. Because of the repetitive nature of this gene family, these sequences tend to be missing from pyrosequenced genome assemblies; therefore, we examined the status of this locus in several the strains via Southern analysis, revealing that this locus is highly plastic both within and between strains (Figure 3).

High density sequence coverage of additional strains

An additional two transmissible strains (Virginia and Puerto Rico) and one non-transmissible strain (Mississippi) were subjected to genome-scale pyrosequencing [24] (454 Life Sciences, Branford, CT), which provided at least 96% genome coverage when compared to either Florida or St. Maries (Table 3). Most of the missing sequences corresponded to repetitive regions (such as *msp2* and *msp3* pseudogenes and the *aaap* locus) (Figure 4), and reflects the limitations of assembling short sequence reads (averaging approximately 250 bp per read) without additional scaffolding. No new genes were detected in the pyrosequenced contigs of any of the strains.

Diversity of *A. marginale* strains

Global comparison of all strains with the Florida strain revealed 20,028 total sites with a single nucleotide polymorphism (SNP) in at least one of the compared strains. Of these, 511 (2.6%) were different in the Florida genome and identical in the other four strains, and 13,316 (66.5%) were unique to one of the four strains (Figure 5). The

remaining 30.9% of SNPs represent those SNPs relative to Florida that were present in two or three of the strains. There were 9,609 SNPs between the Florida and St. Maries strains, comprising 0.80% of the larger Florida genome. The SNPs were distributed evenly throughout the genome, which is similar to both *Ehrlichia ruminantium* and *Rickettsia bellii* (Supplemental figure 1), and are proportionally distributed throughout coding and non-coding regions. The numbers of polymorphisms in the Puerto Rico, Virginia, and Mississippi strains (2,729, 3,868, and 6,773, respectively) are minimums, as the gaps are regions predicted to have significant numbers of SNPs. When the genome size was corrected for the gaps in coverage, the SNP rates for the Puerto Rico, Virginia, and Mississippi genomes were 0.32%, 0.46%, and 0.73% of the Florida genome, respectively.

DISCUSSION

This study illustrates the dangers of drawing universal conclusions when strains are selected based on specific criteria, such as phenotypic differences. No two strains in this study are truly representative of the population as a whole. Additionally, the large number of differences in pair-wise comparisons of any two strains illustrates the difficulty of associating genes with phenotypic differences, and the utility of sequencing multiple strains to increase the power of these associations. While our initial selection of the Florida strain was based on a phenotypic difference – that of tick transmissibility, the selection of subsequent strains (PR, VA, and MS) were made to try and minimize the effect of bias based on that phenotype, as well as select a wider geographic range of isolates to increase interstrain diversity. Interestingly, when the pyrosequenced strains are compared to Florida, there are more high-quality

polymorphisms (identified when four reads, each with at least 20 base pairs flanking the polymorphic site, contain the difference, with at least one read in each direction) between Florida and the Mississippi strain, despite the fact that neither is tick-transmissible by *D. andersoni* [19, 25]. Further, St. Maries appears to be an outlier sequence, as there are at least 6,000 differences between St. Maries and all other sequenced strains.

The level of SNP diversity in these strains coupled with the high degree of gene content conservation also sheds an interesting light on the concept of the “core genome”, described for *Streptococcus agalactiae* [10]. For *S. agalactiae*, approximately 90.5% of genes were considered part of the “core genome”, or constant between strains, and each new strain added additional strain-specific genes to the “pan-genome”. This is contrasted with *Bacillus anthracis*, which had no new strain-specific genes after four strains were compared. The strains of *A. marginale* sequenced here present an interesting data point, as *A. marginale* has not been hypothesized to be a clonal population derived from another organism (as has been postulated for *B. anthracis*), and yet has a closed core genome. The accumulation of large numbers of SNPs might indicate a greater evolutionary distance; however, the closed-core genome could be due to other factors. These could include the isolated nature of the intracellular niche occupied by *A. marginale*, causing the organism to undergo reductive evolution to the point it is approaching the minimal gene complement, or may be, despite our efforts, related to the strains selected for sequencing. However, if this is due to long-term reductive evolution, it calls into question the source of the six split ORFs between the

Florida and St. Maries genomes, as these are thought to be early reductive changes. Another possibility is that transmission of the organism among animals in a relatively restricted geographic area (i.e., within a herd) promotes a relatively clonal population of organisms through isolation in a similar environment.

Analysis of the level of SNP diversity in several bacterial genomes brings into question previous conclusions about the variability of obligate intracellular pathogens. Previous studies [6] have found relatively large numbers of SNPs between intracellular organisms. It was therefore hypothesized that the relatively isolated intracellular niche limits opportunities for genetic exchange and increased numbers of SNPs provides a compensatory mechanism for providing diversity to drive evolution. Our results suggest this is unlikely, as there is no correlation between intracellular, facultative intracellular, and free-living organisms and the level of diversity. With few exceptions, there is a large range in the degree of variability in all the strains compared. Additionally, the organisms with the two highest rates of variability, *Pseudomonas syringae* and *Rhodopseudomonas palustris*, are both free-living. There is also significant variation at the genus and family level. These data suggest that the factors for retention of SNPs leading to bacterial diversity are likely multifactorial and complex.

While the composition of the gene content of the pan-genome is obviously important, this study reveals another characteristic that needs examination: the level of diversity in the pan-genome. The minimum of 20,028 variable sites found among these five genomes is approximately 1.67% of the estimated size of the pan-genome. The

large number of unique SNPs in each strain (24.1% in the St. Maries genome, 6.0% in the Puerto Rico genome, 10.8% in the Virginia genome, and 25.5% in the Mississippi genome) suggests that while *A. marginale* has a closed core genome, the SNP profile of the core genome is moderately "open". When several strains of *Streptococcus agalactiae* (CJB111, COH1, A909, and 515) are compared to the 2603VR strain, 99.18% of the 46,579 total detected SNPs are unique to an individual strain, while zero SNPs are common to all four strains. Similarly, 100% of SNPs between three strains of *Bacillus anthracis* (Ames, Ames Ancestor, and Sterne) and *Mycobacterium tuberculosis* (F11, H37Ra, and H37Rv), 98.8% of SNPs between three strains of *Neisseria meningitides* (FAM18, MC58, and Z2491), and 99.9% of SNPs between four strains of *Chlamydophila pneumoniae* (AR39, CWL029, J138, and TW-183) are unique to one strain. This suggests that these genomes have open SNP profiles regardless of being open or closed-core at the genome level. Further, there is no correlation between SNP diversity and lifestyle, with high levels of variation between strains and within genera, with limited exceptions. However, given that the majority of strains were selected based on phenotypic traits or previous work with each strain, it is unlikely that this represents the true diversity of these organisms. Additionally, the majority of organisms have only two sequenced strains, making analysis of variation within a species impossible to determine. Additional work will be required to build a picture of genomic diversity.

The genome of *A. marginale* is highly recombinogenic, which, in spite of the highly conserved gene content, leads to increased plasticity. There are between five and nine functional *msp2* pseudogenes in the strains examined to date [11, 22, 26, 27],

and these can recombine in whole or in part into the *msp2* expression site (or with each other) to generate new antigenic variants [26, 27]. Symmetrical inversions around the origin are thought to be quite common in bacteria [28] and have been noted in *Anaplasmataceae*, often utilizing repeated genes such as *msp2* to mediate the inversion. These inversions are highlighted by comparisons between *A. marginale* and *Ehrlichia ruminantium* [29] and *Anaplasma phagocytophilum* [30]. Many of these repetitive sequences flank *ori*, as does another duplicated gene, *rho*. While not around the origin, a smaller scale inversion was found between two strains of *A. marginale* flanked by *msp3* pseudogenes close to *ori*. Another highly plastic genomic region is the AAAP locus [23] that appears to be expanding and contracting within and between strains. In addition to changes in gene number, the sequences are highly variable (Table 2). Further research will be needed to determine the significance of these differences, as well as the function of this locus.

CONCLUSIONS

Sequencing of multiple strains of bacteria, as well as sequencing multiple isolates from the same strain, will yield a tremendous amount of information about natural rates of variation in bacterial populations, which in turn will influence our views of bacterial evolution, epidemiology, and vaccine strategies. This study reveals that interstrain SNP diversity does not appear to be influenced by the environmental niche an organism occupies, nor is it generally consistent throughout a specific family or genera. Comparison of multiple strains of *A. marginale* finds few changes at the gene level, while there is robust diversity at the nucleotide level. Finally, multistrain SNP analysis appears to be a more powerful tool for *A. marginale* phylogenetic studies than

genotyping of the major surface proteins [15], and this strategy should be useful for epidemiologic studies of other species of bacteria.

METHODS

Experimental Approval

All animal experiments described in this paper were approved by the Washington State University Institutional Animal Care and Use Committee (IACUC), with approval number 3386.

Strains of *A. marginale* used

The Florida strain [GenBank: CP001079] of *A. marginale* was originally isolated from a pool of blood samples collected from cattle in 1955 [16, 31, 32]. The Mississippi strain [GenBank: ABOP00000000] was isolated from an acute clinical case of anaplasmosis [25, 33]. Both of these strains are virulent, and are not transmissible by *D. andersoni* ticks. The Virginia strain [GenBank: ABOR00000000] was isolated from a cow in Southern Virginia in 1972 [34]. The Puerto Rico strain [GenBank: ABOQ00000000] was received as a frozen stabilate after isolation from cattle in Puerto Rico in 1985 [35, 36]. Both the Virginia and Puerto Rico strains are virulent, and are transmissible by *D. andersoni*. While passage histories are not well documented for these strains, all strains have been passaged multiple times in cattle since isolation. The Florida strain has the longest passage history, being passed continuously since isolation, and the Puerto Rico strain has only been passaged once since coming to our laboratory.

DNA isolation for genome sequencing

Blood stabilates from the Florida, Mississippi, Puerto Rico, and Virginia strains were inoculated into splenectomized calves, which were shown to be free of *A. marginale* infection via competitive enzyme-linked immunoabsorbent assay (cELISA) [37]. Blood samples were taken at peak parasitemia, washed seven times in phosphate buffered saline (137 mM NaCl, 10 mM Phosphate, and 2.7 mM KCl), and centrifuged at 1,500 x g for 10 minutes with the removal of the buffy coat after each spin. Erythrocytes not used immediately were diluted 1:1 in PBS and frozen for later use.

***A. marginale* preparation**

After thawing, lysed erythrocytes were passed over a column containing loosely-packed CF-11 cellulose (Sigma-Aldrich Corporation, St. Louis, MO). The column eluate was washed repeatedly with PBS and centrifuged at 19,000g for 20 minutes until all remaining hemoglobin was removed, leaving a pellet of *A. marginale* initial bodies and erythrocyte membranes.

Bacterial artificial chromosome (BAC) library construction and manipulation.

The bacterial preparation was embedded into 1% agarose blocks (A-9539, Sigma Chemical Co., St. Louis, MO), and cells within the blocks were lysed using proteinase K and SDS [38]. *A. marginale* genomic DNA was partially digested with either *Hind*III or *Mbo*I, size selected on pulse-field gels, ligated into the pBELOBAC11 vector, and electroporated into *Escherichia coli* strain DH10B (Amplicon Express,

Pullman, WA). A total of 3,072 clones (1,536 clones from each restriction enzyme) were arrayed into 384 well plates. The average insert size of the clones was 120 kb.

Genome sequencing.

For the Florida strain, a BAC-based clone-by-clone strategy was adopted. BAC clones were screened using digoxigenin (DIG)-labeled (Roche Applied Science) probes to bovine genomic DNA and several *A. marginale* genes (including *msp1α*, *msp1β*, *msp2*, *msp3*, *msp4*, *msp5*, *dnaK*, *recA*, *groEL*, and *sodB*). Selected clones were end-sequenced and a minimum tiling path was constructed based on comparison with the previously-sequenced St. Maries strain. Sequencing of BACs, assembly of completed sequences, and genome annotation were as described [11].

Genomic DNA from the Mississippi, Puerto Rico, and Virginia strains was extracted from isolated bacteria (prepared as described above) using the Puregene Blood kit (Qiagen Corporation, Valencia, CA). DNA was then sequenced on a Genome Sequencer 20 instrument (454 Life Sciences Corporation, Branford, CT), using a pyrosequencing protocol [24]. The Newbler program was used with its default settings to assemble the sequence and to compare all contigs to the completed Florida and St. Maries genomes, which revealed the location of gaps in coverage. High-quality variations were called when four reads, each with at least 20 base pairs flanking the polymorphic site, contain the difference, with at least one read in each direction. The nucleotide sequences of assembled contigs were compared to the St. Maries genome using BLASTn. Any contigs without hits better than $1e^{-10}$ were then compared to the

bovine whole genome shotgun sequence database, to screen for bovine DNA contamination. Any contigs with no hits to bovine sequence were then compared to the nt database. Large contigs assembled by Newbler were compared to the Florida genome using MUMmer v3.1 [39] after filling gaps in the assembly with the corresponding sequence from the Florida strain.

Genome comparisons.

MUMmer v3.1 was used to compare the completed St. Maries and Florida genomes and the contigs of the Mississippi, Puerto Rico, and Virginia strains, as described [5]. Output from the SNP detection algorithm (show-snps) was processed using custom scripts (written with AutoIt v3.2.2.0) to determine the number of SNPs per ORF. Show-snps output was also processed in Excel (Microsoft Corporation, Redmond, WA) to graph the location of SNPs throughout the genome. FASTA sequences from single-chromosome genomes with multiple strains sequenced were downloaded from Genbank (see supplementary information for the genomes compared, their sizes, and Genbank accession numbers). All strains for a given species were compared to each other using MUMmer 3.1, as described above. The number of SNPs per comparison was then divided by the larger of the two compared genomes to yield the percent SNPs per genome. For species with more than two strains sequenced, all percentages were averaged to give the mean and standard deviation. The phylogenetic tree was inferred using the Maximum Parsimony method [40] of MEGA4 [41] comparing concatenated sequences from *groEL*, *groES*, *atpA*, and *recA*. The bootstrap consensus tree is inferred from 1000 replicates, and branches corresponding to partitions reproduced in

less than 50% bootstrap replicates are collapsed. There were a total of 458 positions in the final dataset, out of which 379 were parsimony informative.

Southern analysis

Genomic DNA from all five strains was digested with *XbaI* and *HindIII* (New England Biolabs Corporation, Ipswich, MA), as these enzymes cut within the conserved flanking genes. Resultant fragments were separated on a 0.8% agarose gel, and subsequently transferred to a charged nylon membrane and crosslinked with a Stratalinker UV apparatus (Stratagene Inc., La Jolla, CA) per the manufacturer's directions. The blots were prehybridized at 42℃ fo r at least two hours in Dig Easy Hyb buffer (Roche Corporation, Indianapolis, IN). Digoxigenin-labeled probes to *aaap* were produced using the PCR DIG Probe Synthesis Kit (Roche Corporation) and hybridized overnight at 42℃ in DIG Easy Hyb buffer. The membr ane was washed three times for 15 minutes in 2xSSC and 0.1% SDS, with the first two washes at room temperature and the third at 65℃. A final wash was performed in 0. 2xSSC and 0.1% SDS at 65℃. Chemiluminescent detection of the probes was performed using the DIG Wash and Block Buffer Kit (Roche Corporation) per the manufacturer's directions.

REFERENCES

1. Mobius P, Luyven G, Hotzel H, Kohler H: **High Genetic diversity among *Mycobacterium avium* subsp. *paratuberculosis* strains of German cattle herds shown by combination of IS900 RFLP analysis and MIRU-VNTR typing**. *J Clin Microbiol* 2008.

2. Tartof SY, Solberg OD, Riley LW: **Genotypic analyses of uropathogenic *Escherichia coli* based on fimH single nucleotide polymorphisms (SNPs)**. *J Med Microbiol* 2007, **56**(Pt 10):1363-1369.

3. Wu T, Chellemi DO, Graham JH, Martin KJ, Rosskopf EN: **Comparison of soil bacterial communities under diverse agricultural land management and crop production practices**. *Microb Ecol* 2008, **55**(2):293-310.

4. Bottos EM, Vincent WF, Greer CW, Whyte LG: **Prokaryotic diversity of arctic ice shelf microbial mats**. *Environ Microbiol* 2008.

5. Fleischmann RD, Alland D, Eisen JA, Carpenter L, White O, Peterson J, DeBoy R, Dodson R, Gwinn M, Haft D *et al*: **Whole-genome comparison of *Mycobacterium tuberculosis* clinical and laboratory strains**. *J Bacteriol* 2002, **184**(19):5479-5490.

6. Tamas I, Klasson L, Canback B, Naslund AK, Eriksson AS, Wernegreen JJ, Sandstrom JP, Moran NA, Andersson SG: **50 million years of genomic stasis in endosymbiotic bacteria**. *Science* 2002, **296**(5577):2376-2379.

7. Carlson JH, Porcella SF, McClarty G, Caldwell HD: **Comparative genomic analysis of *Chlamydia trachomatis* oculotropic and genitotropic strains**. *Infect Immun* 2005, **73**(10):6407-6418.

8. Pearson T, Busch JD, Ravel J, Read TD, Rhoton SD, U'Ren JM, Simonson TS, Kachur SM, Leadem RR, Cardon ML *et al*: **Phylogenetic discovery bias in *Bacillus anthracis* using single-nucleotide polymorphisms from whole-genome sequencing**. *Proc Natl Acad Sci U S A* 2004, **101**(37):13536-13541.

9. Stratilo CW, Lewis CT, Bryden L, Mulvey MR, Bader D: **Single-nucleotide repeat analysis for subtyping *Bacillus anthracis* isolates**. *J Clin Microbiol* 2006, **44**(3):777-782.

10. Tettelin H, Masignani V, Cieslewicz MJ, Donati C, Medini D, Ward NL, Angiuoli SV, Crabtree J, Jones AL, Durkin AS *et al*: **Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome"**. *Proc Natl Acad Sci U S A* 2005, **102**(39):13950-13955.

11. Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, Palmer GH, McGuire TC, Knowles DP, Jr.: **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* 2005, **102**(3):844-849.

12. Dumler JS, Barbet AF, Bekker CP, Dasch GA, Palmer GH, Ray SC, Rikihisa Y, Rurangirwa FR: **Reorganization of genera in the families *Rickettsiaceae* and *Anaplasmataceae* in the order *Rickettsiales*: unification of some species of *Ehrlichia* with *Anaplasma*, *Cowdria* with *Ehrlichia* and *Ehrlichia* with *Neorickettsia*, descriptions of six new species combinations and designation of *Ehrlichia equi* and 'HGE agent' as subjective synonyms of *Ehrlichia phagocytophila***. *Int J Syst Evol Microbiol* 2001, **51**(Pt 6):2145-2165.

13. Allred DR, McGuire TC, Palmer GH, Leib SR, Harkins TM, McElwain TF, Barbet AF: **Molecular basis for surface antigen size polymorphisms and conservation of a neutralization-sensitive epitope in *Anaplasma marginale***. *Proc Natl Acad Sci U S A* 1990, **87**(8):3220-3224.

14. de la Fuente J, Ruybal P, Mtshali MS, Naranjo V, Shuqing L, Mangold AJ, Rodriguez SD, Jimenez R, Vicente J, Moretta R *et al*: **Analysis of world strains of *Anaplasma marginale* using major surface protein 1a repeat sequences**. *Vet Microbiol* 2007, **119**(2-4):382-390.

15. de la Fuente J, Van Den Bussche RA, Garcia-Garcia JC, Rodriguez SD, Garcia MA, Guglielmone AA, Mangold AJ, Friche Passos LM, Barbosa Ribeiro MF, Blouin EF *et al*: **Phylogeography of New World isolates of *Anaplasma marginale* based on major surface protein sequences**. *Vet Microbiol* 2002, **88**(3):275-285.

16. McGuire TC, Palmer GH, Goff WL, Johnson MI, Davis WC: **Common and isolate-restricted antigens of *Anaplasma marginale* detected with monoclonal antibodies**. *Infect Immun* 1984, **45**(3):697-700.

17. Noh SM, Brayton KA, Knowles DP, Agnes JT, Dark MJ, Brown WC, Baszler TV, Palmer GH: **Differential expression and sequence conservation of the *Anaplasma marginale* msp2 gene superfamily outer membrane proteins**. *Infect Immun* 2006, **74**(6):3471-3479.

18. Eriks IS, Stiller D, Goff WL, Panton M, Parish SM, McElwain TF, Palmer GH: **Molecular and biological characterization of a newly isolated *Anaplasma marginale* strain**. *J Vet Diagn Invest* 1994, **6**(4):435-441.

19. Wickwire KB, Kocan KM, Barron SJ, Ewing SA, Smith RD, Hair JA: **Infectivity of three *Anaplasma marginale* isolates for *Dermacentor andersoni***. *Am J Vet Res* 1987, **48**(1):96-99.

20. Frutos R, Viari A, Ferraz C, Morgat A, Eychenie S, Kandassamy Y, Chantal I, Bensaid A, Coissac E, Vachiery N *et al*: **Comparative genomic analysis of three strains of *Ehrlichia ruminantium* reveals an active process of genome size plasticity**. *J Bacteriol* 2006, **188**(7):2533-2542.

21. Ogata H, Audic S, Renesto-Audiffren P, Fournier PE, Barbe V, Samson D, Roux V, Cossart P, Weissenbach J, Claverie JM *et al*: **Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii***. *Science* 2001, **293**(5537):2093-2098.

22. Rodriguez JL, Palmer GH, Knowles DP, Jr., Brayton KA: **Distinctly different msp2 pseudogene repertoires in *Anaplasma marginale* strains that are capable of superinfection**. *Gene* 2005, **361**:127-132.

23. Stich RW, Olah GA, Brayton KA, Brown WC, Fechheimer M, Green-Church K, Jittapalapong S, Kocan KM, McGuire TC, Rurangirwa FR *et al*: **Identification of a novel *Anaplasma marginale* appendage-associated protein that localizes with actin filaments during intraerythrocytic infection**. *Infect Immun* 2004, **72**(12):7257-7264.

24. Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen YJ, Chen Z *et al*: **Genome sequencing in microfabricated high-density picolitre reactors**. *Nature* 2005, **437**(7057):376-380.

25. Ueti MW, Reagan JO, Jr., Knowles DP, Jr., Scoles GA, Shkap V, Palmer GH: **Identification of midgut and salivary glands as specific and distinct barriers to efficient tick-borne transmission of *Anaplasma marginale***. *Infect Immun* 2007, **75**(6):2959-2964.

26. Brayton KA, Knowles DP, McGuire TC, Palmer GH: **Efficient use of a small genome to generate antigenic diversity in tick-borne ehrlichial pathogens**. *Proc Natl Acad Sci U S A* 2001, **98**(7):4130-4135.

27. Brayton KA, Palmer GH, Lundgren A, Yi J, Barbet AF: **Antigenic variation of *Anaplasma marginale* msp2 occurs by combinatorial gene conversion**. *Mol Microbiol* 2002, **43**(5):1151-1159.

28. Eisen JA, Heidelberg JF, White O, Salzberg SL: **Evidence for symmetric chromosomal inversions around the replication origin in bacteria**. *Genome Biol* 2000, **1**(6):RESEARCH0011.

29. Collins NE, Liebenberg J, de Villiers EP, Brayton KA, Louw E, Pretorius A, Faber FE, van Heerden H, Josemans A, van Kleef M *et al*: **The genome of the heartwater agent *Ehrlichia ruminantium* contains multiple tandem repeats of actively variable copy number**. *Proc Natl Acad Sci U S A* 2005, **102**(3):838-843.

30. Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen J, Seshadri R, Ren Q, Wu M, Utterback TR *et al*: **Comparative genomics of emerging human ehrlichiosis agents**. *PLoS Genet* 2006, **2**(2):e21.

31. Ristic M, Sibinovic S, Welter CJ: **An attenuated *Anaplasma marginale* vaccine**. *Proc Annu Meet U S Anim Health Assoc* 1968, **72**:56-69.

32. Ristic M, Carson CA: **Methods of Immunoprophylaxis Against Bovine Anaplasmosis with Emphasis on use of the Attenuated *Anaplasma marginale* Vaccine**. In: *Immunity to Blood Parasites of Animals and Man.* Edited by Miller LH, Pino JA, McKelvey JJJ. New York: Plenum Press; 1977: 151-188.

33. Hidalgo RJ, Palmer GH, Jones EW, Brown JE, Ainsworth AJ: **Infectivity and antigenicity of *Anaplasma marginale* from tick cell culture**. *Am J Vet Res* 1989, **50**(12):2033-2036.

34. Kuttler KL, Winward LD: **Serologic comparisons of 4 *Anaplasma* isolates as measured by the complement-fixation test**. *Vet Microbiol* 1984, **9**(2):181-186.

35. Scoles GA, Ueti MW, Noh SM, Knowles DP, Palmer GH: **Conservation of transmission phenotype of *Anaplasma marginale* (*Rickettsiales: Anaplasmataceae*) strains among *Dermacentor* and *Rhipicephalus* ticks (*Acari: Ixodidae*)**. *J Med Entomol* 2007, **44**(3):484-491.

36. Futse JE, Ueti MW, Knowles DP, Jr., Palmer GH: **Transmission of *Anaplasma marginale* by *Boophilus microplus*: retention of vector competence in the absence of vector-pathogen interaction**. *J Clin Microbiol* 2003, **41**(8):3829-3834.

37. Torioni de Echaide S, Knowles DP, McGuire TC, Palmer GH, Suarez CE, McElwain TF: **Detection of cattle naturally infected with *Anaplasma marginale* in a region of endemicity by nested PCR and a competitive enzyme-linked immunosorbent assay using recombinant major surface protein 5**. *J Clin Microbiol* 1998, **36**(3):777-782.

38. Birren B, Green ED, Klapholz S, Myers RM, Roskams J: **Genome analysis : a laboratory manual**. Plainview, N.Y.: Cold Spring Harbor Laboratory Press; 1997.

39. Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL: **Versatile and open software for comparing large genomes**. *Genome Biol* 2004, **5**(2):R12.

40. Eck RV, Dayhoff MO: **Atlas of protein sequence and structure**. In. Silver Spring, Md.: National Biomedical Research Foundation.; 1966: 5 v.

41. Tamura K, Dudley J, Nei M, Kumar S: **MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0**. *Mol Biol Evol* 2007, **24**(8):1596-1599.

Table 1 - Comparison of the St. Maries and Florida genome features

	Anaplasma marginale St. Maries	*Anaplasma marginale* Florida
Genome Size (bp)	1,197,687	1,202,435
CDS features	949	942
GC content (%)	49.86%	49.86%
Coding density (%)	85.40%	85.50%
Average gene length	1078	1091
rRNA genes	3	3
tRNA genes	37	37
Functional pseudogenes	16	17

Table 2 - Identity between deduced AAAP amino acid sequences from the St. Maries and Florida strains

		St. Maries *aaap*	St. Maries AM879	St. Maries AM880
Florida	*aaap*	45.0%	49.4%	49.6%
Florida	*alp1*	46.1%	73.8%	24.1%
Florida	*alp2*	33.4%	40.3%	58.2%
Florida	*alp3*	11.7%	30.7%	24.6%

Table 3 - Pyrosequencing results for three strains of *Anaplasma marginale*

	Puerto Rico		**Virginia**		**Mississippi**	
	St. Maries	Florida	St. Maries	Florida	St. Maries	Florida
Number of large contigs	75	59	81	70	78	82
Bases in large contigs	1,150,801	1,158,530	1,146,893	1,153,875	1,139,486	1,141,520
% Genome coverage	96.89%	97.00%	96.88%	96.77%	96.34%	96.25%
High quality variations	6,038	2,729	6,613	3,868	6,302	6,773

Table 4 – Pyrosequencing gaps

A listing of genes containing gaps in the pyrosequenced genomes (Puerto Rico [PR], Virginia [VA], and Mississippi [MS]). Black indicates genes completely missing from the assembly, dark grey indicates less than 50% of the gene contained in the assembly, and light grey indicates more than 50% but less than 100% contained in the assembly.

MS		PR	VA	MS
AMF_873	Conserved hypothetical protein			0.00%
AMF_662	Appendage-associated protein-like protei			19.72%
AMF_154	ORF X			24.87%
AMF_875	Outer membrane protein 2			34.70%
AMF_876	outer membrane protein 3			38.89%
AMF_059	ORF X			62.40%
AMF_422	3-demethylubiquinone-9 3-methyltransferase			65.82%
AMF_343	Conserved hypothetical protein			72.16%
AMF_924	Hypothetical protein			75.70%
AMF_571	Hypothetical protein			79.93%

VA+MS		PR	VA	MS
AMF_756	ORF Y		0.00%	70.46%
AMF_755	ORF X		66.18%	66.67%
AMF_866	major surface protein 2 (MSP2)		77.35%	70.04%

PR		PR	VA	MS
AMF_057	ORF X	45.93%		
AMF_262	Hypothetical protein	59.55%		
AMF_090	major surface protein 1B	78.98%		

VA		PR	VA	MS
AMF_504	Hypothetical protein		56.02%	
AMF_503	Hypothetical protein		58.36%	

PR+VA		PR	VA	MS
AMF_738	fructose-1-6-bisphosphate II (glpX)	57.32%	58.05%	
AMF_033	ORF X	60.10%	63.25%	
AMF_919	outer membrane protein 7	80.49%	51.68%	

PR+MS		PR	VA	MS
AMF_801	ORF X	63.05%		69.77%
AMF_660	Appendage-associated protein-like protein	69.57%		48.74%

All Strains		PR	VA	MS
AMF_048	ORF Y	0.00%	0.00%	64.94%
AMF_024	ORF Y	4.76%	1.73%	68.83%
AMF_663	Appendage-associated protein	9.05%	26.54%	10.80%
AMF_735	Hypothetical protein	13.49%	15.25%	63.76%
AMF_536	ORF X	14.65%	0.00%	3.28%
AMF_664	Appendage-associated protein-like protein 3	16.38%	60.31%	0.00%
AMF_659	Appendage-associated protein	17.80%	30.04%	11.93%
AMF_576	Conserved hypothetical protein	20.01%	20.45%	31.92%
AMF_802	major surface protein 3 (MSP3)	21.76%	3.42%	20.99%
AMF_045	transcription termination factor rho (rho)	23.86%	40.60%	22.98%
AMF_535	ORF Y	25.54%	24.68%	22.51%
AMF_642	Conserved hypothetical protein	26.52%	27.13%	23.88%
AMF_135	major surface protein 1B-1	28.79%	59.70%	64.11%
AMF_449	GTP binding protein (hflX)	31.61%	30.36%	26.44%
AMF_509	Hypothetical protein	32.99%	18.95%	44.85%
AMF_577	Conserved hypothetical protein	33.33%	33.49%	28.39%
AMF_946	ORF Y	33.77%	36.80%	21.21%
AMF_505	Hypothetical protein	33.97%	18.88%	39.42%
AMF_758	transcription termination factor RHO	40.14%	41.81%	40.14%
AMF_698	translation elongation factor Tu (tuf)	40.33%	43.25%	41.33%
AMF_186	translation elongation factor Tu (tuf)	40.71%	41.56%	39.95%
AMF_427	O-sialoglycoprotein endopeptidase (gcp)	46.83%	45.44%	43.65%
AMF_098	major surface protein 1B	60.28%	37.87%	48.79%
AMF_418	O-sialoglycoprotein endopeptidase (gcp)	70.87%	71.26%	70.38%
AMF_023	ORF X	71.03%	75.90%	62.05%
AMF_423	translation initiation factor IF-2	75.52%	74.97%	31.12%

Table 5 – Species compared in SNP analysis

A listing of bacterial species compared in the SNP analysis, with Genbank accession numbers for each genome analyzed.

Class	Order	Family	Genus	Species	#	Strains			Size
Alphaproteobacteria	Rickettsiales	Anaplasmataceae	***Anaplasma***	***marginale***	2	Florida			
						St. Maries	CP000030.1	NC_004842.2	1197687
Bacilli	Bacillales	Bacillaceae	***Bacillus***	***anthracis***	3	Ames Ancestor	AE017334.2	NC_007530.2	5227419
						Ames	AE016879.1	NC_003997.3	5227293
						Sterne	AE017225.1	NC_005945.1	5228663
Bacilli	Bacillales	Bacillaceae	***Bacillus***	***cereus***	4	ATCC 10987	AE017194.1	NC_003909.8	5224283
						ATCC 14579	AE016877.1	NC_004722.1	5411809
						E33L	CP000001.1	NC_006274.1	5300915
						subsp. Cytotoxis NVH 391-98	CP000764.1	NC_009674.1	4087024
Bacilli	Bacillales	Bacillaceae	***Bacillus***	***thuringiensis***	2	serovar konkukian str. 97-27	AE017355.1	NC_005957.1	5237682
						str. Al Hakam	CP000485.1	NC_008600.1	5257091
Bacteroidetes	Bacteroidales	Bacteroidaceae	***Bacteroides***	***fragilis***	2	NCTC 9343	CR626927.1	NC_003228.3	5205140
						YCH46	AP006841.1	NC_006347.1	5277274
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	***Buchnera***	***aphidicola***	4	Str. APS	BA000003.2	NC_002528.1	640681
						Str. Bp	AE016826.1	NC_004545.1	615980
						Str. Cc	CP000263.1	NC_008513.1	416380
						Str. Sg	AE013218.1	NC_004061.1	641454
Epsilonproteobacteria	Campylobacterales	Campylobacteraceae	***Campylobacter***	***jejuni***	3	subsp. Jejuni 81-176	CP000538.1	NC_008787.1	1616554
						subsp. Jejuni 81116	CP000814.1	NC_009839.1	1628115
						subsp. Jejuni NCTC 11168	AL111168.1	NC_002163.1	1641481
Chlamydiae	Chlamydiae	Chlamydiales	***Chlamydia***	***trachomatis***	2	A/HAR-13	CP000051.1	NC_007429.1	1044459
						D/UW-3/CX	AE001273.1	NC_000117.1	1042519
Chlamydiae	Chlamydiae	Chlamydiales	***Chlamydia***	***pneumoniae***	4	AR39	AE002161.1	NC_002179.2	1229853
						CWL029	AE001363.1	NC_000922.1	1230230
						J138	BA000008.3	NC_002491.1	1226565
						TW-183	AE009440.1	NC_005043.1	1225935
Clostridia	Clostridia	Clostridiales	***Clostridium***	***botulinum***	3	A str. ATCC 19397	CP000726.1	NC_009697.1	3863450
						A str. ATCC 3502	AM412317.1	NC_009495.1	3886916
						A str. Hall	CP000727.1	NC_009698.1	3760560
Clostridia	Clostridia	Clostridiales	***Clostridium***	***perfringens***	3	ATCC 13124	CP000246.1	NC_008261.1	3256683
						SM101	CP000312.1	NC_008262.1	2897393
						str. 13	BA000016.3	NC_003366.1	3031430
Gammaproteobacteria	Legionellales	Coxiellaceae	***Coxiella***	***burnetii***	2	Dugway 7E9-12	CP000733.1	NC_009727.1	2158758
						RSA 493	AE016828.2	NC_002971.3	1995281
Deltaproteobacteria	Desulfovibrionales	Desulfovibrionaceae	***Desulfovibrio***	***vulgaris***	2	subsp. Vulgaris DP4	CP000527.1	NC_008751.1	3462887
						subsp. Vulgaris str. Hildenborough	AE017285.1	NC_002937.3	3570858
Alphaproteobacteria	Rickettsiales	Anaplasmataceae	***Ehrlichia***	***ruminantium***	3	Gardel	CR925677.1	NC_006831.1	1499920
						Welgevonden (SA)	CR767821.1	NC_005295.2	1516355
						Welgevonden (French)	CR925678.1	NC_006832.1	1512977
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	***Escherichia***	***coli***	9	536	CP000247.1	NC_008253.1	4938920
						APEC O1	CP000468.1	NC_008563.1	5082025
						CFT073	AE014075.1	NC_004431.1	5231428
						E24377A	CP000800.1	NC_009801.1	4979619
						HS	CP000802.1	NC_009800.1	4643538
						K12	U00096.2	NC_000913.2	4639675
						O157:H7 EDL933	AE005174.2	NC_002655.2	5528445
						O157:H7 str. Sakai	BA000007.2	NC_002695.1	5498450
						UTI89	CP000243.1	NC_007946.1	5065741
Gammaproteobacteria	Thiotrichales	Francisellaceae	***Francisella***	***tularensis***	3	subsp. Holarctica LVS	AM233362.1	NC_007880.1	1895994
						subsp. Holarctica FTA	CP000803.1	NC_009749.1	1890909
						subsp. Holarctica OSU18	CP000437.1	NC_008369.1	1895727
Gammaproteobacteria	Thiotrichales	Francisellaceae	***Francisella***	***tularensis***	3	subsp. Tularensis FSC198	AM286280.1	NC_008245.1	1892616
						subsp. Tularensis SCHU S4	AJ749949.1	NC_006570.1	1892819
						subsp. Tularensis WY96-3418	CP000608.1	NC_009257.1	1898476
Gammaproteobacteria	Pasteurellales	Pasteurellaceae	***Haemophilus***	***influenzae***	4	86-028NP	CP000057.2	NC_007146.1	1914490
						PittEE	CP000671.1	NC_009566.1	1813033
						PittGG	CP000672.1	NC_009567.1	1887192
						Rd KW20	L42023.1	NC_000907.1	1830138
Epsilonproteobacteria	Campylobacterales	Helicobacteraceae	***Helicobacter***	***pylori***	3	26695	AE000511.1	NC_000915.1	1667867
						HPAG1	CP000241.1	NC_008086.1	1596366
						J99	AE001439.1	NC_000921.1	1643831
Gammaproteobacteria	Legionellales	Legionellaceae	***Legionella***	***pneumophila***	4	str. Corby	CP000675.1	NC_009494.1	3576470
						str. Lens	CR628337.1	NC_006369.1	3345687
						str. Paris	CR628336.1	NC_006368.1	3503610
						sbsp. Pneumophila str. Philadelphia 1	AE017354.1	NC_002942.5	3397754

Bacilli	Bacillales	Listeriaceae	***Listeria***	***monocytogenes***	2	EGD-e	AL591824.1	NC_003210.1	2944528
						str. 4b F2365	AE017262.2	NC_002973.6	2905187
Actinobacteria	Actinomycetales	Mycobacteriaceae	***Mycobacterium***	***bovis***	2	AF2122/97	BX248333.1	NC_002945.3	4345492
						BCG str. Pasteur 1173P2	AM408590.1	NC_008769.1	4374522
Actinobacteria	Actinomycetales	Mycobacteriaceae	***Mycobacterium***	***tuberculosis***	4	CDC1551	AE000516.2	NC_002755.2	4403837
						F11	CP000717.1	NC_009565.1	4424435
						H37Ra	CP000611.1	NC_009525.1	4419977
						H37Rv	AL123456.2	NC_000962.2	4411532
Mollicutes	Mycoplasmatales	Mycoplasmataceae	***Mycoplasma***	***hyopneumoniae***	3	232	AE017332.1	NC_006360.1	892758
						7448	AE017244.1	NC_007332.1	920079
						J	AE017243.1	NC_007295.1	897405
Betaproteobacteria	Neisseriales	Neisseriaceae	***Nesseria***	***meningitidis***	3	FAM18	AM421808.1	NC_008767.1	2194961
						MC58	AE002098.2	NC_003112.2	2272360
						Z2491	AL157959.1	NC_003116.1	2184406
Cyanobacteria	Prochlorales	Prochlorococcaceae	***Prochlorococcus***	***marinus***	9	AS9601	CP000551.1	NC_008816.1	1669886
						MIT 9215	CP000825.1	NC_009840.1	1738790
						MIT 9301	CP000576.1	NC_009091.1	1641879
						MIT 9303	CP000554.1	NC_008820.1	2682675
						MIT 9312	CP000111.1	NC_007577.1	1709204
						MIT 9313	BX548175.1	NC_005071.1	2410873
						MIT 9515	CP000552.1	NC_008817.1	1704176
						NATL1A	CP000553.1	NC_008819.1	1864731
						NATL2A	CP000095.2	NC_007335.1	1842899
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	***Pseudomonas***	***aeruginosa***	3	PA7	CP000744.1	NC_009656.1	6588339
						PAO1	AE004091.2	NC_002516.2	6264404
						UCBPP-PA14	CP000438.1	NC_008463.1	6537648
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	***Pseudomonas***	***fluorescens***	2	Pf-5	CP000076.1	NC_004129.6	7074893
						PfO-1	CP000094.1	NC_007492.1	6438405
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	***Pseudomonas***	***putida***	2	F1	CP000712.1	NC_009512.1	5959964
						KT2440	AE015451.1	NC_002947.3	6181863
Gammaproteobacteria	Pseudomonadales	Pseudomonadaceae	***Pseudomonas***	***syringae***	3	pv. Phaseolicola 1448A	CP000058.1	NC_005773.3	5928787
						pv. Syringae B728a	CP000075.1	NC_007005.1	6093698
						pv. Tomato str. DC3000	AE016853.1	NC_004578.1	6397126
Alphaproteobacteria	Rhizobiales	Bradyrhizobiaceae	***Rhodopseudomor***	***palustris***	5	BisA53	CP000463.1	NC_008435.1	5505494
						BisB18	CP000301.1	NC_007925.1	5513844
						BisB5	CP000283.1	NC_007958.1	4892717
						CGA009	BX571963.1	NC_005296.1	5459213
						HaA2	CP000250.1	NC_007778.1	5331656
Alphaproteobacteria	Rickettsiales	Rickettsiaceae	***Rickettsia***	***bellii***	2	RML 369-C	CP000087.1	NC_007940.1	1522076
						OSU 85-389	CP000849.1	NC_009883.1	1528980
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	***Salmonella***	***enterica***	2	subsp. Enterica serovar typhi Ty2	AE014613.1	NC_004631.1	4791961
						subsp. Enterica serovar typhi str. CT18	AL513382.1	NC_003198.1	4809037
Gammaproteobacteria	Alteromonadales	Shewanellaceae	***Shewanella***	***baltica***	2	OS155	CP000563.1	NC_009052.1	5127376
						OS185	CP000753.1	NC_009665.1	5229686
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	***Shigella***	***flexneri***	2	2a str. 2457T	AE014073.1	NC_004741.1	4599354
						2a str. 301	AE005674.1	NC_004337.1	4607203
Bacilli	Bacillales	Staphylococcaceae	***Staphylococcus***	***aureus***	13	RF122	AJ938182.1	NC_007622.1	2742531
						subsp. Aureus COL	CP000046.1	NC_002951.2	2809422
						subsp. Aureus JH1	CP000736.1	NC_009632.1	2906507
						subsp. Aureus JH9	CP000703.1	NC_009487.1	2906700
						subsp. Aureus MSRA252	BX571856.1	NC_002952.2	2902619
						subsp. Aureus MSRA476	BX571857.1	NC_002953.3	2799802
						subsp. Aureus MW2	BA000033.2	NC_003923.1	2820462
						subsp. Aureus Mu3	AP009324.1	NC_009782.1	2880168
						subsp. Aureus Mu50	BA000017.4	NC_002758.2	2878529
						subsp. Aureus N315	BA000018.3	NC_002745.2	2814816
						subsp. Aureus NCTC 8325	CP000253.1	NC_007795.1	2821361
						subsp. Aureus USA300	CP000255.1	NC_007793.1	2872769
						subsp. Aureus str. Newman	AP009351.1	NC_009641.1	2878897
Bacilli	Bacillales	Staphylococcaceae	***Staphylococcus***	***epidermidis***	2	ATCC 12228	AE015929.1	NC_004461.1	2499279
						RP62A	CP000029.1	NC_002976.3	2616530
	Chroococcales		***Synechococcus***	***elongatus***	2	PCC 6301	AP008231.1	NC_006576.1	2696255
						PCC 7942	CP000100.1	NC_007604.1	2695903
Deinococci	Thermales	Thermaceae	***Thermus***	***thermophilus***	2	HB27	AE017221.1	NC_005835.1	1894877
						HB8	AP008226.1	NC_006461.1	1849742
Actinobacteria	Actinomycetales	Cellulomonadaceae	***Tropheryma***	***whipplei***	2	TW08/27	BX072543.1	NC_004551.1	925938
						Twist	AE014184.1	NC_004572.3	927303

Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	***Xanthomonas***	***campestris***	3	pv. Campestris str. 8004	CP000050.1	NC_007086.1	5148708
						pv. Campestris str. ATCC 33913	AE008922.1	NC_003902.1	5076188
						pv. Campestris str. 85-10	AM039952.1	NC_007508.1	5178466
Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	***Xanthomonas***	***oryzae***	2	pv. Oryzae KACC10331	AE013598.1	NC_006834.1	4941439
						pv. Oryzae MAFF 311018	AP008229.1	NC_007705.1	4940217
Gammaproteobacteria	Xanthomonadales	Xanthomonadaceae	***Xylella***	***fastidiosa***	2	9a5c	AE003849.1	NC_002488.3	2679306
						Temecula1	AE009442.1	NC_004556.1	2519802
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	***Yersinia***	***pestis***	6	Antiqua	CP000308.1	NC_008150.1	4702289
						CO92	AL590842.1	NC_003143.1	4653728
						KIM	AE009952.1	NC_004088.1	4600755
						Nepal516	CP000305.1	NC_008149.1	4534590
						Pestoides F	CP000668.1	NC_009381.1	4517345
						biovar. Microtus str. 91001	AE017042.1	NC_005810.1	4595065
Gammaproteobacteria	Enterobacteriales	Enterobacteriaceae	***Yersinia***	***pseudotuberculo:***	2	IP 31758	CP000720.1	NC_009708.1	4723306
						IP 32953	BX936398.1	NC_006155.1	4744671

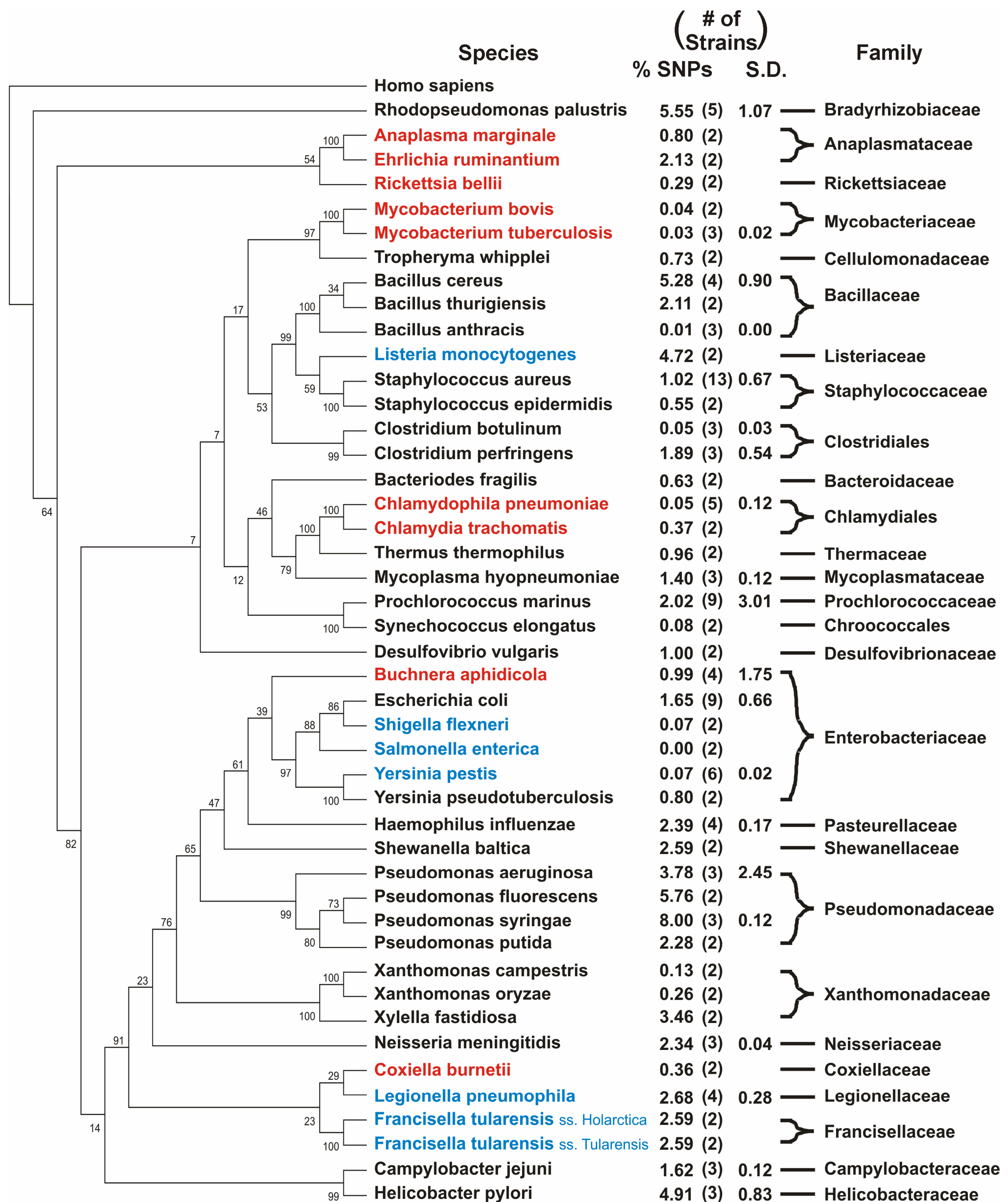

Figure 1 - Comparison of the level of SNP diversity among sequenced genomes. Bacteria listed in blue are obligate intracellular, while those in red are facultatively intracellular. Each organism lists the average level of SNPs as a % of the largest genome.

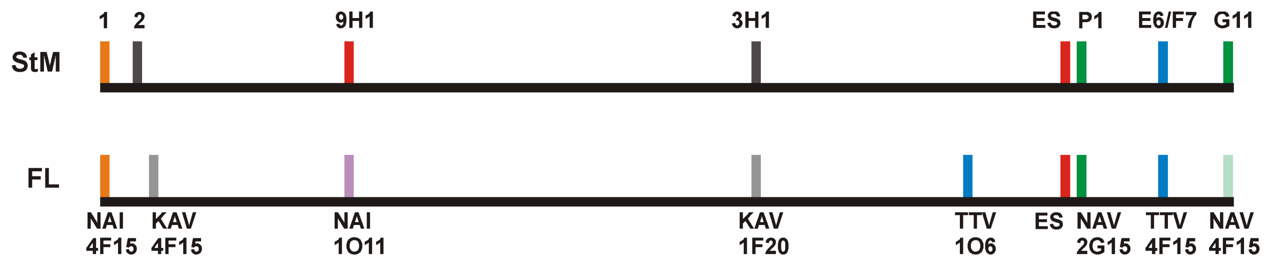

Figure 2 - Physical map of the MSP2 functional pseudogenes in the St. Maries (StM) and Florida (FL) strains.
Lines (not to scale) with the same color indicate identical pseudogenes, while similar colors indicate pseudogenes with segmental changes. ES represents the msp2 expression site.

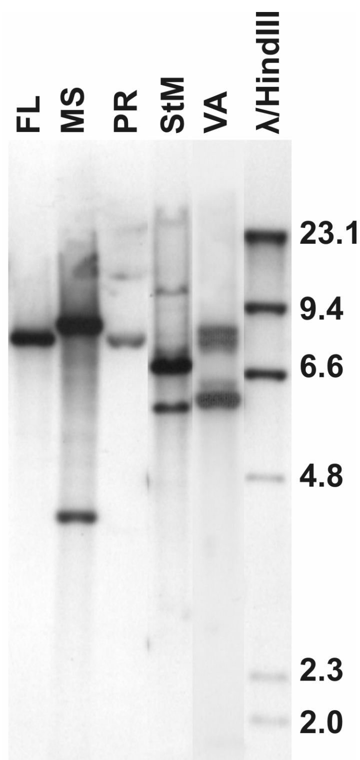

Figure 3 - Southern blot of the aaap locus in *A. marginale* strains.
FL - Florida strain, MS - Mississippi strain, PR - Puerto Rico strain, StM - St. Maries strain, VA - Virginia strain. DNA marker sizes are listed in kbp.

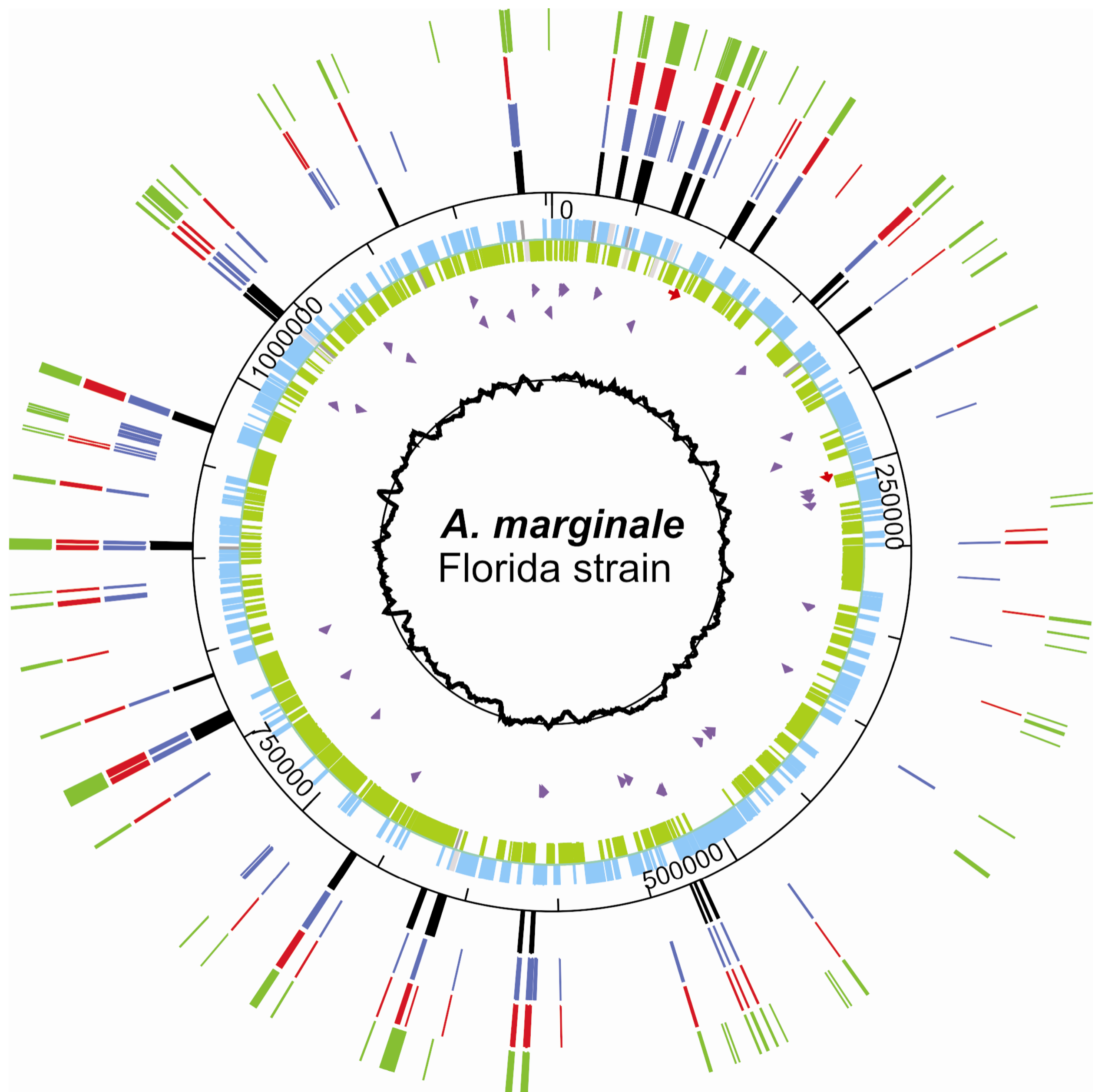

Figure 4 - Distribution of gaps in the three pyrosequenced genomes.
In the outer rings, gaps in the Mississippi sequence are green, gaps in Puerto Rico are red, and gaps in Virginia are blue. Known repetitive genes are represented by black bars. The inner rings represent CDSs (blue and green) and functional pseudogenes (shades of grey) in the Florida strain, rRNAs (red) and tRNAs (purple), and the G-C skew (black graph).

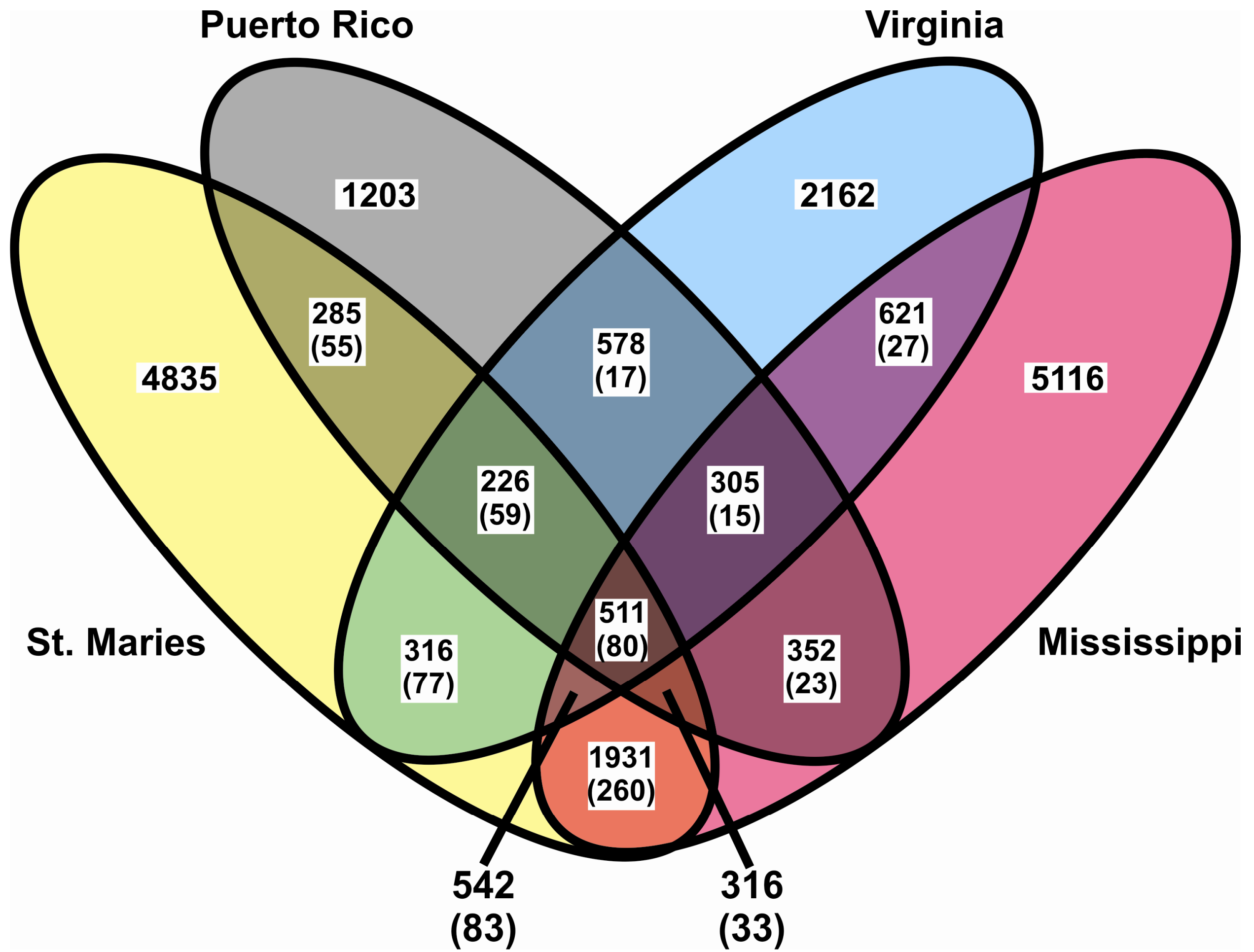

Figure 5 - Distribution of SNPs in four strains compared to the Florida strain. Numbers in parentheses show SNPs that are different from Florida in each strain in the subset, but are not the same in the compared strains.

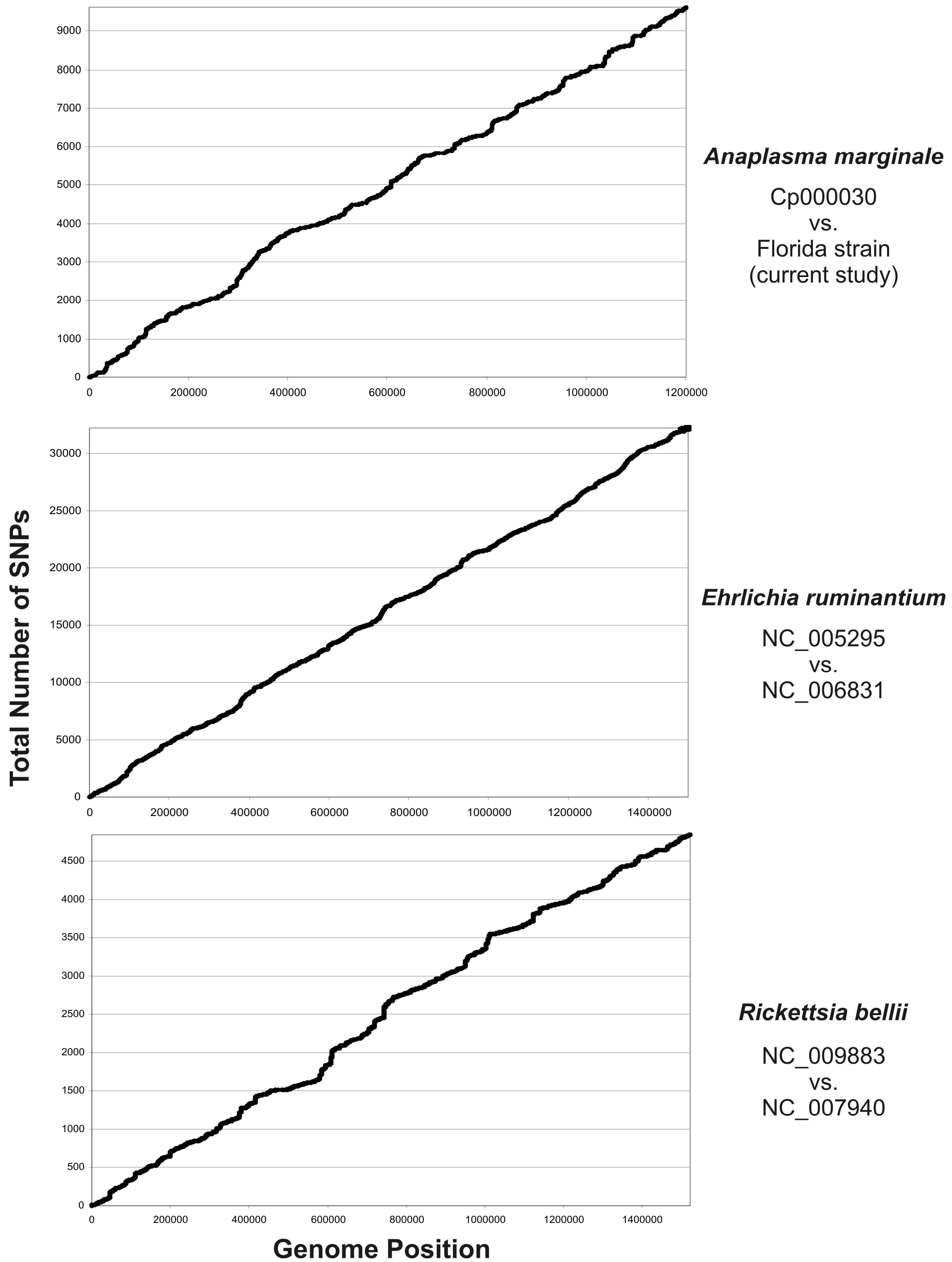

Figure 6 – SNP distribution in three species
Distribution of SNPs between the compared strains of *A. marginale*, *E. ruminantium*, and *R. bellii*.

CHAPTER 2

EXAMINATION OF TWO STRAINS OF *ANAPLASMA MARGINALE* REVEALS DIFFERENTIAL VECTOR COMPETENCE WITH *DERMACENTOR ANDERSONI* AND *RHIPICEPHALUS MICROPLUS*

ABSTRACT

Since the discovery of strains of *Anaplasma marginale* that are not transmissible by tick vectors in 1966, there has been considerable interest in determining the factors involved in vector infection. While several theories have been proposed, none have proved fruitful. The most recent hypothesis, proposed in 2001, is that the N-terminal repeats of the Msp1a molecule effect tick transmission via binding to tick epithelial cells. Additionally, previous studies have demonstrated that different strains of *Anaplasma marginale* retain similar abilities to infect vectors, despite geographic separation. To test both of these hypotheses, we infected calves with the South Idaho and Mississippi strains of *A. marginale*, which have identical Msp1a repeats (DDDDE). At peak parasitemia, adult male *Dermacentor andersoni* and *Rhipicephalus* (*Boophilus*) *microplus* ticks were fed to repletion and subsequently placed on susceptible calves. The South Idaho strain was transmitted by *D. andersoni*, but not *R. microplus*, while the Mississippi strain was not transmitted by either tick. This demonstrates that the sequence differences of the Msp1a repeats are not involved in tick transmission, and that not all strains retain the ability to be transmitted by all vectors.

INTRODUCTION

The requirement for vectored transmission of infectious pathogens, such as Babesia, Plasmodium, Borrelia, African trypanosomes, Ehrlichia and Anaplasma is necessary for epidemiological persistence. Elucidation of the nature of the interaction between pathogen and vector is crucial to the development of interventions to prevent the spread of disease, such as transmission blocking vaccines. To study the interaction between pathogen and vector we have utilized the model organism *Anaplasma marginale*, the most prevalent tick-transmitted pathogen of cattle (13). *A. marginale* is biologically transmitted between hosts via ixodid ticks. The availability of strains of *A. marginale* that differ in tick transmission status allows investigation into the complex vector-pathogen relationship.

While the most common tick vector of *A. marginale* in the northwestern United States is *Dermacentor andersoni*, in the areas of the world where *A. marginale* has the highest prevalence, *Rhipicephalus* (*Boophilus*) *microplus* is the more common vector (8). Although *R. microplus* is not found in the United States, it is capable of transmitting strains of *A. marginale* found in Idaho, suggesting that vector competence is retained in the absence of pathogen-vector interactions (8).

While no genetic differences have been proven to cause differences in transmission status, one study (4) implicated Msp1 as the reason for this phenotypic difference. Msp1 is formed by a complex of Msp1a and Msp1b molecules (14). While no specific function has been determined for this protein, expression of Msp1a allows for adhesion of *Escherichia coli* to cultured tick cells (3). When adhesion of Msp1a molecules from different strains are compared, *E. coli* expressing the Florida strain molecule have

significantly less adhesion to tick cells than *E. coli* expressing Msp1a from the tick-transmissible Oklahoma strain (4). As the major difference between the Msp1a molecules resides in the number and sequence of repeats at the amino terminus of the molecule, the repeat region was hypothesized to be responsible for the presence or lack of binding to tick cells and thus lack of transmission status observed in the Florida strain (7, 21).

To test this hypothesis, we examined the ability of two strains, South Idaho (11) and Mississippi (9, 20), to be transmitted by tick vectors. These strains have identical Msp1a repeat regions, providing an opportunity to test whether the Msp1a repeat structure is involved in tick transmission. At the same time, we tested the ability of *R. microplus* ticks to transmit both of these strains, to determine if the defect in the Mississippi strain was restricted to *D. andersoni*, and to determine if geographically similar tick-transmitted strains (St. Maries and South Idaho) retain competence across the range of tick vectors for this organism.

MATERIALS AND METHODS

Strains Used

The South Idaho strain was isolated from a clinical case in Caldwell, ID in 1983 (11). The Mississippi strain has been stored as a stabilate in liquid nitrogen since being used in laboratory experiments in the early 1980s (9).

Transmission

All animal experiments described herein were approved by the Washington State University Institutional Animal Care and Use Committee (IACUC), with approval number 3386. Ten age-matched Holstein calves were determined to be free of *A. marginale* infection via competitive enzyme-linked immunoabsorbent assay (cELISA) (19) (VMRD, Pullman, WA). One calf was inoculated with the South Idaho strain and one inoculated with the Mississippi strain by intravenous injection of 10^9 organisms in a blood stabilate. Subsequent infection was evaluated using Giemsa-stained blood smears and cELISA testing. After *A. marginale* organisms were detected in erythrocytes, adult male *Dermacentor andersoni* (Reynolds Creek strain) and *Rhipicephalus microplus* (La Minita strain) (17) were allowed to feed under separate cloth patches for seven days. Ticks were removed, incubated at 27℃ for two days, and p laced in groups of 30 onto two susceptible calves for each species of tick for transmission feeding. After seven days, the ticks were removed and ten ticks per calf were dissected, with midguts and salivary glands collected from each dissected tick for confirmation of infection via PCR (see

below). Calf infection status was followed using blood smears, cELISA, and PCR as previously described.

DNA isolation

Blood samples taken at peak parasitemia were washed seven times in phosphate buffered saline (PBS; 137 mM NaCl, 10 mM Phosphate, and 2.7 mM KCl), and centrifuged at 1,500 x g for 10 minutes with the removal of the buffy coat after each spin. Erythrocytes not used immediately were diluted 1:1 in PBS and frozen for later use. Genomic DNA was extracted from initial bodies using the Puregene Blood kit (Qiagen Corporation, Valencia, CA), and DNA was extracted from dissected tick salivary glands and midguts as previously described (10, 17), and resuspended in 30µL of DNA hydration solution (Qiagen Corportation, Valencia, CA).

Genotyping

Msp1α forward primer (5' ATT TCC ATA TAC TGT GCA G) and reverse primer (5' CTT GGA GCG CAT CTC TCT TGC C) were used to generate amplicons from genomic DNA as previously described (1, 12). PCR products were purified using the QIAquick PCR Purification Kit (Qiagen Corporation, Valencia, CA) and sequenced using the forward and reverse primers with BigDye Terminator v3.1 Cycle Sequencing Kit (Applied Biosystems Inc., Foster City, CA) on an Applied Biosystems 3130xl Genetic Analyzer. DNA sequences were analyzed with Sequencher (Gene Codes, Inc., Ann Arbor, MI) and deduced amino acid sequences were compared to published MSP1a repeat sequences (1, 6).

Primers AMF_530_F2 (5' GAC GTC CAA CCA GGA AAA CTC ATA TGA C 3') and AMF_530_R2 (5' CTT TTC GTA TGC GTC TTG CG 3') were used to amplify a polymorphic gene in South Idaho and Mississippi using PCR (95℃ for 2 minutes, followed by 40 cycles of 95℃ for 30 seconds, 53℃ for 45 seconds, and 72℃ for 2 minutes, followed by a final extension of 72℃ for 7 minutes). The amplicon in Mississippi contains an extra HpaII restriction site, allowing differentiation of South Idaho and Mississippi via HpaII (New England Biolabs, Ipswich, MA) digestion of amplicons, with visualization of the resulting fragments on a 1% agarose gel.

Infection Confirmation

Infection status of tick midguts and salivary glands was evaluated by Southern analysis for the *msp5* gene (8). Briefly, 1µL from each midgut and salivary gland DNA sample was used as the template in a PCR reaction for *msp5*, using the primers MSP5-367F (5'-TAC ACG TGC CCT ACC GAG TTA-3') and MSP5-710R (5'-TCC TCG CCT TGG CCC TCA GA-3') and the following reaction conditions: 96℃ for 2 minutes, 40 cycles of 96℃ for 30 seconds, 58℃ for 30 seconds, and 72℃ for 1 minute, and a final 72℃ extension for 7 minutes. Resultant fragments w ere separated on a 1.0% agarose gel, transferred to a charged nylon membrane, and crosslinked with a Stratalinker UV apparatus (Stratagene Inc., La Jolla, CA) per the manufacturer's directions. The blots were prehybridized at 42℃ for at least two hours i n Dig Easy Hyb buffer (Roche Corporation, Indianapolis, IN). A digoxigenin-labeled *msp5* probe was produced using the PCR DIG Probe Synthesis Kit (Roche Corporation) and hybridized to the membrane

overnight at 42℃ in DIG Easy Hyb buffer. The membrane was washed three times for 15 minutes in 2xSSC and 0.1% SDS, with the first two washes at room temperature and the third at 65℃. A final wash was performed in 0.2xSSC and 0.1% SDS at 65℃. Chemiluminescent detection of the probes was performed using the DIG Wash and Block Buffer Kit (Roche Corporation) according to the manufacturer's directions.

For quantitative PCR, 20µL reactions were prepared as previously described (17), using 5'- CCT CCG CGT CTT TCA ACA ATT TGG TT -3' as the probe.

RESULTS

Genotype of A. marginale strains

A. marginale strains are typically distinguished by *msp1α* genotype (5). Both the South Idaho and Mississippi strains have the MSP1a repeat sequence DDDDE (per the scheme described in 1). This is contrasted with the original characterization of the South Idaho strain as DDDDDE (1). As both strains in this study have identical *msp1α* genotypes, a second typing method was developed based on analysis of single nucleotide polymorphisms (SNPs) from recently sequenced strains (Dark et. al., submitted). We found a SNP in a gene encoding a hypothetical protein (corresponding to AM712 in the completely sequenced St. Maries strain (2)) that distinguishes the South Idaho strain from the Mississippi strain by creation of an additional HpaII restriction enzyme site in the Mississippi strain.

Acquisition rates and levels of A. marginale in tick midguts and salivary glands

Long acquisition feed lengths (7 days) were used to favor acquisition of *A. marginale* in both the tick midgut and salivary gland. *A. marginale* specific PCR and Southern analysis was utilized to determine the infection rate of ticks as this technique can detect as few as 50 organisms per salivary gland pair (data not shown). *D. andersoni* ticks (80-90%) had significantly higher infection rates than *R. microplus* ticks (10-25%) in both the midgut and salivary gland, regardless of infecting strain (Table 1). However, quantitative analysis reveals that although a high proportion of *D. andersoni* ticks were infected with the Mississippi strain, the level of infection was significantly lower (2-3 logs) than *D. andersoni* infected with the South Idaho strain. Similarly, *R. microplus* ticks uniformly had low levels if infection, irrespective of infecting strain (Figure 1). There is no significant difference between the infection levels of the *R. microplus* ticks infected with either strain, and between the two species of ticks infected with the Mississippi strain. At the time the ticks were removed from the calves, the infection level of the calf infected with the South Idaho strain was 3% parasitized erythrocytes (PPE), and the Mississippi level was 3.6% PPE. Therefore, the infection rates in ticks were not merely a reflection of the infection level in the calf at the time of tick feeding.

Transmission of A. marginale strains to naïve animals

Similar to acquisition, long transmission feed times (7 days) were used to favor a transmission event. Despite similar acquisition rates for both strains, *D. andersoni* transmitted only the South Idaho strain to naïve animals. These animals tested positive for South Idaho strain infection two weeks after transmission feeding, using the

genotype specific SNP test described above. *R. microplus* did not transmit either strain (Table 2). Animals classified as negative did not develop infection during the 75-day postfeeding observation period and did not seroconvert.

DISCUSSION

While a number of Rickettsial species are transmitted by tick vectors, no pathogen genes have been confirmed to be involved in tick transmission. Several studies have attempted to define *A. marginale* factors that differentiate tick transmissible and non-transmissible strains (4, 18, 21), with the most recent hypothesis being that the repeat structure of Msp1a is responsible for tick transmission (4). Our results clearly show that, in side by side comparisons, the South Idaho strain was transmissible by the Reynolds Creek stock of *D. andersoni*, while the Mississippi strain was not. As the South Idaho and Mississippi strains have identical Msp1a repeat sequences, we therefore reject the hypothesis that the Msp1a repeat structure alone is responsible for the ability of an *A. marginale* strain to be transmitted. However, there is a caveat to this conclusion: the ability to be transmitted could be a multi-factorial process, with different factors being altered for the different non-transmissible strains. The possibility exists that the Florida strain is not transmissible due to the Msp1a repeat structure (ABBBBBBB) of that strain, while the genetic basis for the lack of tick transmission of the Mississippi strain lies within another gene.

Several previous studies of *A. marginale* transmission have identified strains that appear not to be tick transmissible (9, 15, 20, 21), however, it is not known if this simply

reflects an incorrect pairing of the pathogen and tick vector. For this reason we employed two species of ticks, *D. andersoni* and *R. microplus* that have each been shown to be competent vectors of *A. marginale* (8). This is the first report of differential transmission status for a strain of *A. marginale*, indicating that vector competence phenotypes are more strain-restricted than previously thought. The data suggest that transmission is linked to infection level within the tick, as ticks infected with an average of 150 organisms per salivary gland pair or less failed to transmit, even when the *A. marginale* strain employed has been demonstrated to be competent for transmission. These findings agree with that of Ueti and coworkers, who, working with the vaccine strain of *A. marginale* (previously thought not to be tick transmissible), have achieved transmission by increasing the number of ticks and therefore the total pathogen burden on the transmission animal to that seen with more highly transmissible strains when fewer ticks are applied (Ueti et al., in press). Therefore, the "non-transmissible" Mississippi strain detailed in this study may simply require a higher tick burden for transmission than necessary for highly vector-competent strains. The readily transmissible St. Maries strain is reproducibly capable of transmission with as few as 10 ticks, and has reported salivary gland infection levels of 10^4 - 10^6 (8, 10, 16). We used three times more ticks to bias in favor of transmission; however, if the infection level in the salivary gland is the strongest indicator of transmission, we would need approximately 6-10 times more ticks to achieve transmission with the Mississippi strain. The differential infection levels/transmission status found with the South Idaho strain in the two species of ticks suggests complex interactions that are not solely governed by having the necessary tools for transmission encoded by the pathogen; the tick vector

must also have the complementary machinery for this intricate interaction to result in successful transmission.

The interactions between pathogens and their vectors are one of the least understood areas of infectious disease biology. This study provides an examination of Msp1a and functional testing of its role in tick transmission, as well as the first description of a differential transmission phenotype between different vectors. This phenotypic difference can now be leveraged to examine the required interactions between rickettsial organisms and their vectors, as well as determine factors required of the pathogen and vector for disease transmission.

REFERENCES

1. **Allred, D. R., T. C. McGuire, G. H. Palmer, S. R. Leib, T. M. Harkins, T. F. McElwain, and A. F. Barbet.** 1990. Molecular basis for surface antigen size polymorphisms and conservation of a neutralization-sensitive epitope in *Anaplasma marginale*. Proc Natl Acad Sci U S A **87:**3220-3224.
2. **Brayton, K. A., L. S. Kappmeyer, D. R. Herndon, M. J. Dark, D. L. Tibbals, G. H. Palmer, T. C. McGuire, and D. P. Knowles, Jr.** 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.
3. **de la Fuente, J., J. C. Garcia-Garcia, E. F. Blouin, and K. M. Kocan.** 2001. Differential adhesion of major surface proteins 1a and 1b of the ehrlichial cattle pathogen *Anaplasma marginale* to bovine erythrocytes and tick cells. Int J Parasitol **31:**145-153.
4. **de la Fuente, J., J. C. Garcia-Garcia, E. F. Blouin, B. R. McEwen, D. Clawson, and K. M. Kocan.** 2001. Major surface protein 1a effects tick infection and transmission of *Anaplasma marginale*. Int J Parasitol **31:**1705-1714.
5. **de la Fuente, J., P. Ruybal, M. S. Mtshali, V. Naranjo, L. Shuqing, A. J. Mangold, S. D. Rodriguez, R. Jimenez, J. Vicente, R. Moretta, A. Torina, C. Almazan, P. M. Mbati, S. T. de Echaide, M. Farber, R. Rosario-Cruz, C. Gortazar, and K. M. Kocan.** 2007. Analysis of world strains of *Anaplasma marginale* using major surface protein 1a repeat sequences. Vet Microbiol **119:**382-390.
6. **de la Fuente, J., R. A. Van Den Bussche, and K. M. Kocan.** 2001. Molecular phylogeny and biogeography of North American isolates of Anaplasma marginale (Rickettsiaceae: Ehrlichieae). Vet Parasitol **97:**65-76.

7. **Friedhoff, K. T., and M. Ristic.** 1966. Anaplasmosis. XIX. A preliminary study of *Anaplasma marginale* in *Dermacentor andersoni* (Stiles) by fluorescent antibody technique. Am J Vet Res **27:**643-646.

8. **Futse, J. E., M. W. Ueti, D. P. Knowles, Jr., and G. H. Palmer.** 2003. Transmission of *Anaplasma marginale* by *Boophilus microplus*: retention of vector competence in the absence of vector-pathogen interaction. J Clin Microbiol **41:**3829-3834.

9. **Hidalgo, R. J., G. H. Palmer, E. W. Jones, J. E. Brown, and A. J. Ainsworth.** 1989. Infectivity and antigenicity of *Anaplasma marginale* from tick cell culture. Am J Vet Res **50:**2033-2036.

10. **Lohr, C. V., F. R. Rurangirwa, T. F. McElwain, D. Stiller, and G. H. Palmer.** 2002. Specific expression of Anaplasma marginale major surface protein 2 salivary gland variants occurs in the midgut and is an early event during tick transmission. Infect Immun **70:**114-120.

11. **McGuire, T. C., G. H. Palmer, W. L. Goff, M. I. Johnson, and W. C. Davis.** 1984. Common and isolate-restricted antigens of *Anaplasma marginale* detected with monoclonal antibodies. Infect Immun **45:**697-700.

12. **Palmer, G. H., A. F. Barbet, A. J. Musoke, J. M. Katende, F. Rurangirwa, V. Shkap, E. Pipano, W. C. Davis, and T. C. McGuire.** 1988. Recognition of conserved surface protein epitopes on Anaplasma centrale and Anaplasma marginale isolates from Israel, Kenya and the United States. Int J Parasitol **18:**33-38.

13. **Palmer, G. H., and T. F. McElwain.** 1995. Molecular basis for vaccine development against anaplasmosis and babesiosis. Vet Parasitol **57:**233-253.

14. **Palmer, G. H., F. R. Rurangirwa, K. M. Kocan, and W. C. Brown.** 1999. Molecular basis for vaccine development against the ehrlichial pathogen *Anaplasma marginale*. Parasitol Today **15:**281-286.

15. **Ristic, M., S. Sibinovic, and C. J. Welter.** 1968. An attenuated *Anaplasma marginale* vaccine. Proc Annu Meet U S Anim Health Assoc **72:**56-69.

16. **Scoles, G. A., J. A. Miller, and L. D. Foil.** 2008. Comparison of the efficiency of biological transmission of Anaplasma marginale (Rickettsiales: Anaplasmataceae) by Dermacentor andersoni Stiles (Acari: Ixodidae) with mechanical transmission by the horse fly, Tabanus fuscicostatus Hine (Diptera: Muscidae). J Med Entomol **45:**109-114.

17. **Scoles, G. A., M. W. Ueti, S. M. Noh, D. P. Knowles, and G. H. Palmer.** 2007. Conservation of transmission phenotype of *Anaplasma marginale* (*Rickettsiales: Anaplasmataceae*) strains among *Dermacentor* and *Rhipicephalus* ticks (*Acari: Ixodidae*). J Med Entomol **44:**484-491.

18. **Stich, R. W., G. A. Olah, K. A. Brayton, W. C. Brown, M. Fechheimer, K. Green-Church, S. Jittapalapong, K. M. Kocan, T. C. McGuire, F. R. Rurangirwa, and G. H. Palmer.** 2004. Identification of a novel *Anaplasma marginale* appendage-associated protein that localizes with actin filaments during intraerythrocytic infection. Infect Immun **72:**7257-7264.

19. **Torioni de Echaide, S., D. P. Knowles, T. C. McGuire, G. H. Palmer, C. E. Suarez, and T. F. McElwain.** 1998. Detection of cattle naturally infected with *Anaplasma marginale* in a region of endemicity by nested PCR and a competitive enzyme-linked immunosorbent assay using recombinant major surface protein 5. J Clin Microbiol **36:**777-782.

20. **Ueti, M. W., J. O. Reagan, Jr., D. P. Knowles, Jr., G. A. Scoles, V. Shkap, and G. H. Palmer.** 2007. Identification of midgut and salivary glands as specific and distinct barriers to efficient tick-borne transmission of *Anaplasma marginale*. Infect Immun **75:**2959-2964.

21. **Wickwire, K. B., K. M. Kocan, S. J. Barron, S. A. Ewing, R. D. Smith, and J. A. Hair.** 1987. Infectivity of three *Anaplasma marginale* isolates for *Dermacentor andersoni*. Am J Vet Res **48:**96-99.

Table 1. Analysis of salivary gland and midgut infection status via msp5 Southern blotting after aquisition feeding

		Calf A				Calf B				Total			
		SG		MG		SG		MG		SG		MG	
SI	*D. andersoni*	10	100%	10	100%	6	60%	8	80%	16	80%	18	90%
	B. microplus	3	30%	3	30%	1	10%	2	20%	4	20%	5	25%
MS	*D. andersoni*	8	80%	7	70%	10	100%	9	90%	18	90%	16	80%
	B. microplus	0	0%	1	10%	3	30%	1	10%	3	15%	2	10%

Table 2. Transmission feeding results.

		Calf A	Calf B
South Idaho	*Dermacentor andersoni*	+	+
	Rhipicephalus microplus	-	-
Mississippi	*Dermacentor andersoni*	-	-
	Rhipicephalus microplus	-	-
+: Positive by PCR, blood smear, and cELISA			
-: Negative by PCR, blood smear, and cELISA			

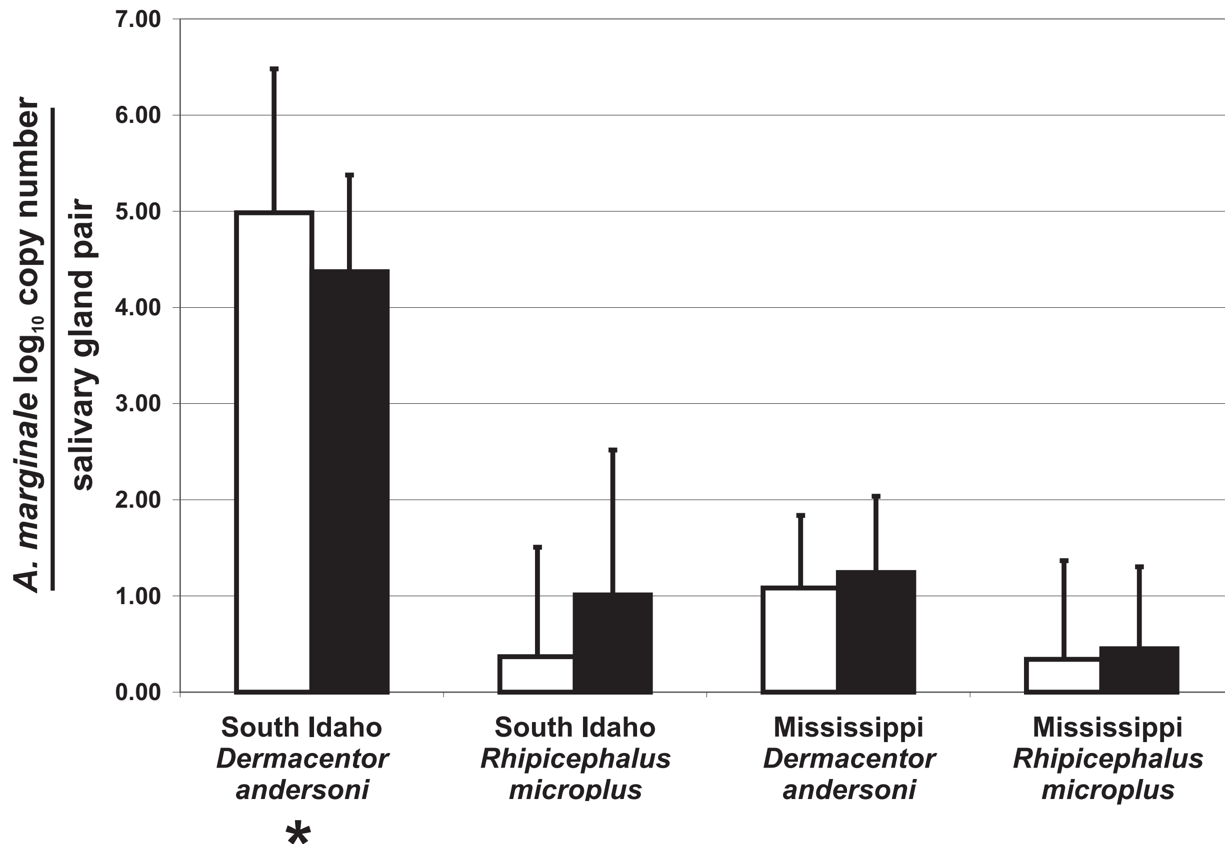

Figure 1: *A. marginale* infection levels in tick tissues. White bars represent levels in salivary gland pairs, and black bars represent midguts. Infection levels were determined using quantitative PCR for *msp5*. The asterisk indicates statistically significant differences, with a $p < 0.01$. Error bars represent one standard deviation, with tests done in triplicate with ten biological replicates. Samples that were below the detection level were assigned the lowest detectable level (50 organisms), to increase stringency of statistical analysis.