PLOS ONE

Functional and Immunological Relevance of *Anaplasma marginale* Major Surface Protein 1a Sequence and Structural Analysis

Alejandro Cabezas-Cruz[1], Lygia M. F. Passos[2,3], Katarzyna Lis[2], Rachel Kenneil[2], James J. Valdés[1], Joana Ferrolho[4], Miray Tonk[1], Anna E. Pohl[2], Libor Grubhoffer[1], Erich Zweygarth[2], Varda Shkap[5], Mucio F. B. Ribeiro[6], Agustín Estrada-Peña[7], Katherine M. Kocan[8], José de la Fuente[8,9]*

1 University of South Bohemia, Faculty of Science and Biology Centre of the Academy of Sciences of the Czech Republic, Parasitology Institute, České Budějovice, Czech Republic, 2 Institute for Comparative Tropical Medicine and Parasitology, Ludwig-Maximilians-Universität, München, Germany, 3 Departamento de Medicina Veterinária Preventiva, Escola de Veterinária, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, 4 Pirbright Laboratory, The Pirbright Institute, Pirbright, Surrey, United Kingdom, 5 Division of Parasitology, Kimron Veterinary Institute, Bet Dagan, Israel, 6 Departamento de Parasitologia, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil, 7 Facultad de Veterinaria, Universidad de Zaragoza, Zaragoza, Spain, 8 Department of Veterinary Pathobiology, Center for Veterinary Health Sciences, Oklahoma State University, Stillwatert, Oklahoma, United States of America, 9 Sanidad y Biotecnología, Instituto de Investigación de Recursos Cinegéticos, IREC-CSIC-UCLM-JCCM, Ciudad Real, Spain

Abstract

Bovine anaplasmosis is caused by cattle infection with the tick-borne bacterium, *Anaplasma marginale*. The major surface protein 1a (MSP1a) has been used as a genetic marker for identifying *A. marginale* strains based on N-terminal tandem repeats and a 5′-UTR microsatellite located in the *msp1a* gene. The MSP1a tandem repeats contain immune relevant elements and functional domains that bind to bovine erythrocytes and tick cells, thus providing information about the evolution of host-pathogen and vector-pathogen interactions. Here we propose one nomenclature for A. marginale strain classification based on MSP1a. All tandem repeats among *A. marginale* strains were classified and the amino acid variability/frequency in each position was determined. The sequence variation at immunodominant B cell epitopes was determined and the secondary (2D) structure of the tandem repeats was modeled. A total of 224 different strains of *A. marginale* were classified, showing 11 genotypes based on the 5′-UTR microsatellite and 193 different tandem repeats with high amino acid variability per position. Our results showed phylogenetic correlation between MSP1a sequence, secondary structure, B-cell epitope composition and tick transmissibility of *A. marginale* strains. The analysis of MSP1a sequences provides relevant information about the biology of *A. marginale* to design vaccines with a cross-protective capacity based on MSP1a B-cell epitopes.

Introduction

Bovine anaplasmosis, caused by the intraerythrocytic rickettsia *Anaplasma marginale* (Rickettsiales: Anaplasmataceae), is an economically important disease of cattle which is endemic in tropical and subtropical regions of the world [1,2]. This obligate intracellular pathogen can be transmitted biologically by ticks, mechanically by transfer of infective blood on fomites or the mouthparts of biting insects [1,2], and, less commonly, by transplacental transmission from dams to their calves [3].

Many geographic strains of *A. marginale* have been identified worldwide which differ in morphology, protein sequence, antigenic characteristics and their ability to be transmitted by ticks [1,2,4–15]. The genetic diversity of *A. marginale* strains derived from bovine erythrocytes has been characterized based on the sequence of major surface protein (MSP) genes, several of which

have been shown to be involved in host cell/pathogen interactions [16]. MSP1a, one of six MSPs described previously on *A. marginale*, is a 70–100 kDa protein encoded by a single-copy gene, *msp1a*, which is conserved during the multiplication in cattle and ticks [17]. MSP1a is involved in adhesion of *A. marginale* to bovine erythrocytes and tick cells and therefore is a determinant of infection for cattle and transmission of *A. marginale* by ticks. MSP1a has also been shown to be involved in development of bovine immunity against *A. marginale* [3]. Strains of *A. marginale* were originally identified by differences in the molecular weight of MSP1a because of variable number of 23–31 amino acid serine-rich tandem repeats located in the N-terminal region of the protein which is continuous with a highly conserved C-terminal region [6,11,14]. Because the number and sequence of tandem repeats remained the same in a given strain, the *msp1a* gene was

recognized as a stable genetic marker for geographic strain identity [9,12,15,18–20]. Phylogenetic analyses of *A. marginale* strains using MSPs were reported by de la Fuente et al. [14,21–23]. While sequence analysis of MSP4 provided phylogeographic information, MSP1a did not prove to be as suitable for these studies [24]. However, MSP1a repeat sequence analysis contributed to the understanding of the genetic diversity of *A. marginale* within specific regions, as well providing insight into the evolution of host–pathogen-vector interactions [14,21–23,25].

MSP1a also contains neutralization sensitive T- and B-cell epitopes required for development of a protective immune response [8,10,26–29]. One B-cell epitope within the MSP1a tandem repeat ((Q/E)ASTSS) was recognized by a monoclonal antibody that neutralized *A. marginale* in vitro [6]. This neutralization-sensitive epitope was found to be conserved among heterologous *A. marginale* strains [29,30]. An additional linear B-cell epitope (SSAGGQQQESS) was found to be immuno dominant [26,28,31]. Cattle immunized with MSP1 were partially protected against challenge with homologous and heterologous strains [32–34]. Furthermore, MSP1a antibodies reduced the infectivity of *A. marginale* for cultured tick cells [35] and infection and transmission of *A. marginale* by *D. variabilis* [1].

MSP1a is relevant to many facets of *A. marginale* research. Strain classification enables a comprehensive study of the extensive worldwide diversity of *A. marginale*. As reported herein, development of an unified nomenclature of MSP1a from *A. marginale* strains based on all available sequence data allowed for review and characterization of the worldwide genetic diversity of *A. marginale*. The information generated from these studies will be fundamental toward understanding the functional and immune relevance of *A. marginale* MSP1a and in formulating vaccines that will be cross-protective among these diverse strains.

Results and Discussion

Classification of A. marginale Strains Using MSP1a Sequence Data

In this study we propose a unified nomenclature for the classification of *A. marginale* strains based on the sequences of the MSP1a tandem repeats and the 5′-UTR microsatellite. This approach was supported by the following considerations: (i) the availability of numerous *A. marginale* MSP1a sequences in GenBank, (ii) the fact that MSP1a is encoded by a single-copy gene [1], (iii) the tandem repeat structure and sequence vary among strains from different geographic locations, while the remaining portion of the protein is highly conserved [14], (iv) the tandem repeats structure is a stable genetic marker that is conserved within a strain during the acute and persistent chronic phases of the *A. marginale* infection in cattle and after passage and transmission by ticks [1], (v) the tandem repeats contain functional domains that serve as adhesins for bovine erythrocytes and tick cells, a prerequisite for infection of host cells [10,36], (vi) the tandem repeats contain relevant B cell epitopes and neutralization epitopes important for natural or induced immune protection in cattle [6,31], and (vii) a microsatellite which has been implicated in the regulation of MSP1a expression levels is located in the 5′-UTR of the *msp1a* gene [25].

In this study, 193 different MSP1a tandem repeats were identified, 79 of which were published in GenBank but not formally classified (Fig. 1). Two new microsatellite structures were described in our analysis and named J and K (J: m = 1, n = 8, d = 21; K: m = 2, n = 8, d = 25) after Estrada-Peña et al. [25]. Unique *A. marginale* strains (224; 77% of all sequences found) are based on differences in geographic location, the number and

structure of the MSP1a tandem repeats and microsatellites when available. These *A. marginale* strains came from 17 world regions providing a global MSP1a diversity (Fig. 2), and were classified following our proposed nomenclature (Table S1). The majority of *A. marginale* strains had more than one MSP1a tandem repeat and the maximum number of repeats was 10. No strains were reported with 9 tandem repeats (Table S1 and Fig. 3). Tables 1 provide a list of the most commonly reported strains and tandem repeats. The majority of strains were seen in only a given region, although several strains were isolated from multiple South American countries (Argentina/Chaco/− (τ, 22, 13, 18) from Argentina and Mexico; Brazil/Parana/− (τ, 10, 15) from Brazil and Argentina; Mexico/Pichucalco/E - (α, β, β, Γ²) from Argentina, Brazil and Mexico; and Mexico/Tamaulipas/− (64, 65, D, 65, E) from Mexico and Venezuela). The strain, Argentina/Santa Fe/− (α, β³, Γ), was the only strain found in more than one continent, and was reported in Argentina, Mexico, and Taiwan. Most of the MSP1a tandem repeats were shared between different strains, and repeat B, the most common tandem repeat sequence, occurred in 43 strains (Table 1). While some tandem repeats were unique to one country (repeat 72 was only reported in Brazil) or continent (repeat B was found throughout the American continent), some repeats appeared to be distributed worldwide (repeat M was reported in Israel, Italy, USA and South America). This weak association between specific tandem repeat sequences and particular geographic regions was reported previously by de la Fuente et al. [14] and may be attributed to worldwide cattle movement, among other factors. Notably, in Australia, in which introduction of cattle has been limited, only one MSP1a genotype has been reported [37].

The Biological Implications of Sequence Variation of MSP1a Tandem Repeats

The tandem repeated portion of the N-terminal region of the *A. marginale* MSP1a has been shown to be an adhesin for bovine erythrocytes and tick cells, and thus are involved in pathogen infection of host cells and transmission by ticks [10,36,38]. In contrast, the MSP1a N-terminal tandem repeats are absent in *A. marginale* subsp. centrale. Although *A. centrale* can be transmitted by *Rhipicephalus simus*, the tick species from which this organisms was initially isolated, this *Anaplasma* sp. cannot be transmitted by other tick species that are known to be *A. marginale* vectors [20,39].

These analyses provided information on the range and frequency of variations in the *A. marginale* MSP1a tandem repeats. Herein, we present the sequence variation data and discuss biological implications of these findings, including O-glycosylation, amino acids at position 20 for binding to tick cell extract (TCE), protein conformation, pathogen-environmental relationships, and combination of these factors.

O-glycosylation. MSP1a tandem repeats were found to have a high variability across almost all the 31 amino acid positions, suggesting considerable evolutionary pressure on this molecule (Fig. 4A). Four positions were totally conserved: serine (S)4 and S25, alanine (A)22 and Glicine 31 (Fig. 4A). MSP1a has been shown to be O-glycosylated, with S/threonine (T) regions present in the tandem repeats as the target site for this type of glycosylation [31]. Furthermore, the binding capacity of MSP1a to tick cells diminished after deglycosylation [31]. The conservation of S4 and S25 among all the tandem repeats included in this study could indicate that the O-glycosylation at these two positions is highly relevant for *A. marginale* infection. Several bacterial glycoproteins have also been reported to play a role in bacterial adhesion, invasion and pathogenesis [40].

```
A   DDSSSASGQQQESSVSSQSE-ASTSS-QLG--      38  A............L...GQ.....-.S.--    107  A......D.....G.P....-....-...--
B   A.....G...........DQ.....-...--        39  ............L...DQ.....-...--     108  A.......------..L...GQ....P-...--
C   A.....G.........GQ.....-...--          40  AG...GD.........DQ.....-...--     109  A........P....L.P.GQ.....-...--
D   A..........-.....-...G-                41  AS...........L...DQ.....-...--    110  A........R..G.L...GQ.....--H..--
E   A..........-.....-...--                42  T...........LP..GQ.....-.S.--     111  T.......R......L...DQ.....-.SR--
F   T...........GQ.....-...--              43  A...........LP..GQ.....-.S.--     112  T...........L...DQ.C.....-...--
G   ...........GQ.....-.S.--               44  T......A.........GQ.....-.S.--    113  T.......R...L...DQ.....-...--
H   T...........GQ.....-.S.--              45  T...........LP..DQ.....-...--     114  T.....GDH..G.G.....GE.....-...--
I   ...........GQ.....-...--               46  T...........LP..GQ.....-...--     115  T.....GDH..G.G.....GQ.....-...--
J   A...L.G.........DQ.....-...--          47  A....GD.........DQ.....-...--     116  T...........T.....D-.....-...--
K   A.G...G.........DQ.....-...--          48  T.G...........L.P.GQ.....-...VG   117  AG....G.........D-.A...YL..--
L   AG....D.........DQ.....-...--          49  T.G...GD...G.G.....GQ.....-...--  118  A.................D-PA...PL..--
M   A...........GQ.....-...--              50  T.....GDR..G.G.....GQ.....-...--  119  A...........L...D-PA....SL..--
m   A...........GQ.....-.S.--              51  T.....GD.R.G.G.....GQ.....-...--  120  T...........DD.TS..-..S--
N   T...........DQ.....-...--              52  T...........V.....D-.....-...--   121  T.....H.R..T.L.R.DQ.....-...--
O   ----..G.........DQ.....-...--          53  T...........LPP.GQ.....-...--     122  T.....H.R..T.L...DQ.....-..S.--
P   T.........G....GQ..H.A-.S.--           54  LH...........F.TGQ.....-...--     123  T.....R..T.L...DQ.....--HS.--
Q   A...........DQ.....-...--              55  T.....GD.H.........GQ.....-...--  124  T.....R..T.L...DQ.....-...--
R   A...G...H.......DQ.....-..W--          56  T.......E.....L.P.GE.....-...--   125  T.......H...L...DQ.....-..S.--
T   AG....G.........DQ.....-...--          57  T.........H...I.....-.......W--   126  T...........N.L.PCGR.....-...--
U   ...........DQ.....-...--               58  T.G...........L...DE.....-...--   127  A.....GD..K..N....TGQ..P..-H.R--
V   -AD..SA.G.........DQ.....-...--        59  T.G...........L...DE.....-.S.--   128  T.....GD...G.G..A..GQ..PW..-...--
W   T...........GQ.....-.SR--              60  A.G...GD.........G-.....-...--    129  A.....RH.K.N.......-.C...-...--
α   A.......------..L...GQ.....-...--      61  T.....GD.........G-.....-...--    130  T.........WC.......-.....-...--
β   T...GD...G.G.....GQ.....-...--         62  T.....GD.........D-.....-...--    131  T...................-....-..R--
γ   T...............D-.....-...--          63  AG...........L...GQ.....-...--    132  A.....G..Q...V.S.DD.G...-P.WT--
π   A....G...........GQ.....-.F.--         64  AG...G...........-.....-...G-     133  -.D.....G..Q..A.P.GD.A...-P.LT--
Σ   A.....G.........-.....-...--           65  A...........L...-.....-...G-      134  -AD..S..D..QN..A.P.DD.A...-H.WT--
σ   A....G....I.....DH.....-...--          66  T.....GD.........-.....-..R--     135  -.D..ST.D..Q...V.P.DD.AR..-..WT--
μ   A......L.........-.....-...--          67  T...........L...GQ..P..-...--     136  A...................-....--RG-
τ   T...........L.P.GQ.....-...--          68  AGR...........L...GQ.....-...--   137  A...........I......-.R....--G-
φ   T...............-.....-...--           69  A.G...GD.........G-.P.....-...--  138  A...............-..S..-..G-
1   SG...........L..GGQ.....-...--         70  T.....GD.....A....G-.....-...--   139  A...........I...........--G-
2   T.............P..GQ.....-...--         71  T.....GD.H.........G-.R...--      140  A...........G.L...GQ.....-..R--
3   A...........L...GQ.....-...--          72  A....GD..-.........G-.....-...--  141  T...........G.L...GQ.....-..R--
4   T...........L...GQ.....-...--          73  T.....GD....G.....GQ.....-...--   142  T...........LP..GH.R...-.S.--
5   A...............D-.....-...--          74  A...........G.P...GQ.....-...--   143  T....G.........L...SQ.RS.-....--
6   A...........H..-.....-...--            75  A.G...GD........G.L...GQ.....-...--  144  .............LP..GQD.....-.S.--
7   T...........H..-.....-...--            76  T...........G....D-.....-...--    145  G....S.........L...SQ.....-...--
8   A.G...GD.........G-.....-.S.--         77  A....V.........L...GQ.....-...--  146  A....GN.............-....-...--
9   A...............D-.....-.S.--          78  T...........L...SQ.....-...--     147  T.....GN.........G-.G....-...--
10  A...........L.P.GQ.....-...--          79  T.......P.....LCV.DLS....-...--   148  T.....GD.........G-..A...-K.R--
11  A...........L.P.GQ.....-..VG           80  A....GD.........GQ.....-...--     149  T.....GD.........L.G-.....-...--
12  AG...........L...DQ.....-...--         81  T.....GD.........GQ.....-...--    150  T.....GD.K.........IG-.......K..--
13  T...........L...DQ.....-...--          82  A...........L...DLS..W..-...--    151  AN........E.....L...DQ.....-...--
14  T...........L..G-.....-...--           83  T...........G.L...GQ.....-..S.--  152  T...........L...DQ.....-..R--
15  A.........G.L...GQ.....-...--          84  A....GN.........G-.....-...--     153  T...PEM.........F...AQ.....-.S.--
16  A....GD...G....GQ.....-...--           85  A.G...GD.........G-.GP..-...--    154  A...........L...DQ.....-.S.--
17  T...........G...GQ.....-...--          86  T.....GD.........K.G-.....-...--  155  AN..........L...GQ.....-...--
18  T...........L...DQ.....-.S.--          87  A...........L...IQ.....-.R--      156  A....GN.........GH.....-...--
19  A....GDR...G.L...GQ.....-...--         88  T.......K...LP..GQ.....-...--     157  T..W.........L...DQ.....-...--
20  A....GD...G.L...GQ.....-...--          89  A....G.........E....DQ.....--K..--  158  AE.........L...D-Q.....-...--
21  A....GD....L...GQ.....-...--           90  A...........DQ..S...-..--         159  A.............P....-.A...-..---
22  A...........L.P.GQ.....-.S.--          91  T...........GQ.....-..R--         160  ANG.........L...DQ.....-...--
23  T.......K...L...SQ.....-...--          92  A.........G.L...TQ.....-.R--      161  A...........L...DE.....-...--
24  A....GN....LP..GQ.....-.S.--           93  N...........L.P.GQ.....-..VG
25  A...........L...SQ.....-...--          94  A.........G.L...GQ.....-..VG
26  A....GN....LP..GQ.....-...--           95  A....TG.........H..-.....-..G-
27  A...........L...DQ.....-...--          96  A...........L....-....-..H..G-
28  AG...E.........L...GQ.....-...--       97  A....P.............-.A...-EVRG-
29  T....D.........GQ.....-...--           98  A...........L.....-.....-..--
30  A.......K...L...SQ.....-...--          99  A..WP..NR...P.L.H.Y-TT..-.RG-
31  A....GN.........D-.....-...--          100  T...........G.L...GQ.....-...--
32  T....G.........GQ.....-...--           101  AG...G.........DQ.T....-...--
33  A...........L.K.GQ.G...-...--          102  A...G.........DQ.N...-H..--
34  AN...........DQ.....-...--             103  A.........G...L...GQ.....-...--
35  T...........GQ.....-.S.--              104  A....E.........L...GQ.....-...--
36  A...........P....-.....-...--          105  T....G.........D-.....-...--
37  T...........L..GQ.....-.S.--           106  A.G...GD.........GQ.....-...--
```

Figure 1. MSP1a tandem repeat sequences in A. marginale strains. The one letter amino acid code was used to depict MSP1a repeat sequences. Dots indicate identical amino acids and gaps indicate deletion/insertions. The ID of each repeat form was given following the nomenclature proposed by de la Fuente et al. (2007) [14]. The sequences from 114 until 161 are the newly classified.
doi:10.1371/journal.pone.0065243.g001

Relevance of amino acids at position 20 for binding to tick cell extract (TCE). Within the MSP1a tandem repeats, the negatively charged amino acids, aspartic acid (D) and glutamic acid (E), at position 20 were shown to be essential for binding of MSP1a to TCE. When glycine (G) was located at position 20, binding was not observed [10]. This result suggested that the amino acid at position 20 may be essential for *A. marginale* binding to tick cells, a prerequisite for pathogen infection and transmission by ticks. In fact, previous experiments confirmed the existence of both tick-transmissible and not transmissible *A. marginale* strains and, at least for some strains, the presence of TCE-binding with tandem repeats correlated with strains that were transmissible by *Dermacentor sp.* ticks [10]. In all strains, the first MSP1a tandem repeat (R1) contained 67 (34.7%) different sequences. However, R1 tandem repeats had less amino acid variability and 6 conserved positions when compared to non-R1 tandem repeats, in which only 4 conserved amino acid positions were found (Fig. 4B). These results suggested that the R1 tandem repeat may play a role in *A.*
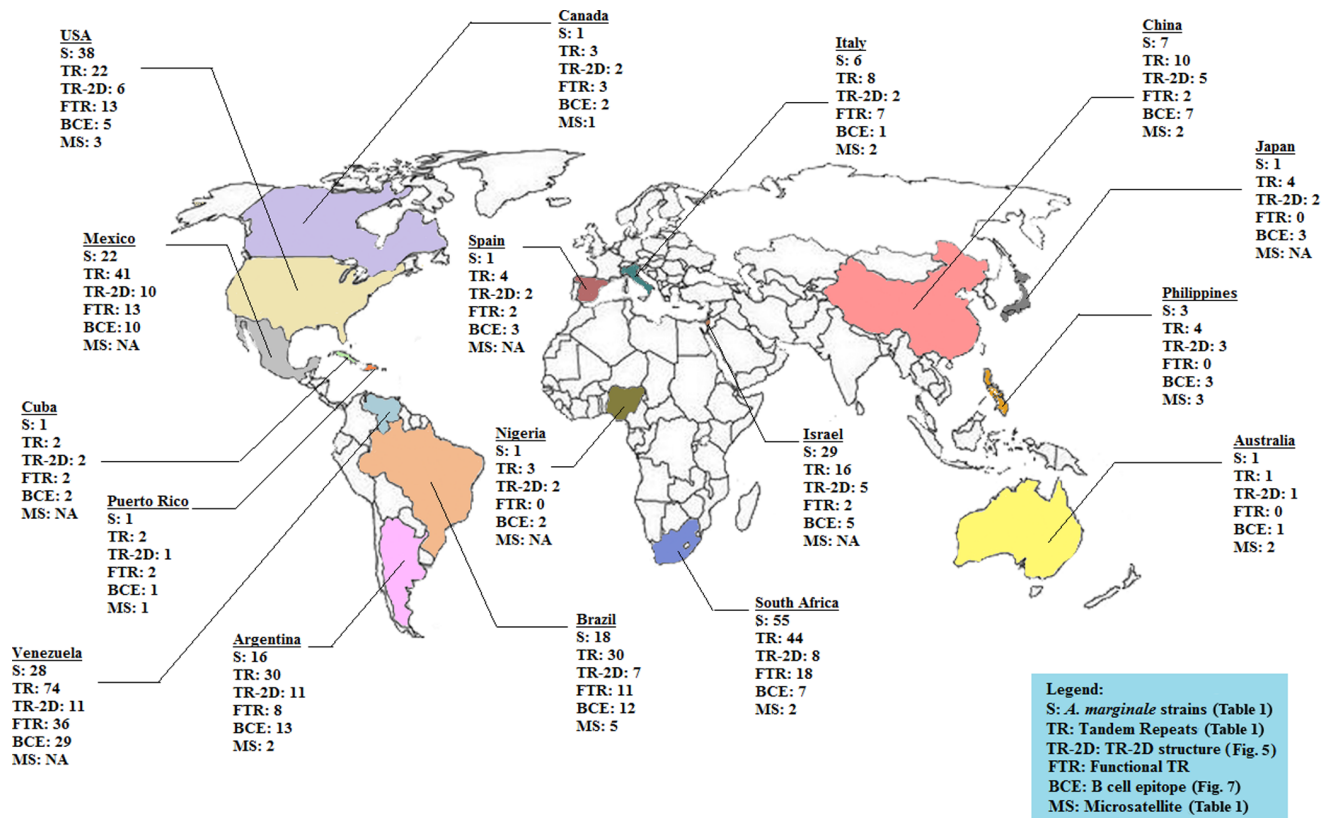
Figure 2. World A. marginale MSP1a molecular map. The worldwide molecular characterization of A. marginale MSP1a sequences is shown. The number of A. marginale strains (S), tandem repeats (TR), tandem repeat 2D structures (TR-2D), functional tandem repeats (FTR) containing D and E at position 20 and B cell epitope types (BCE) and microsatellites (MS) are represented for each country. Primary data is depicted in figures 1, 3 and 6. The information on 5′ UTR microsatellites is not available (NA) for some sequences.

doi:10.1371/journal.pone.0065243.g002

marginale infection and transmission. We found 87 tandem repeats containing D20 (71%) or E20 (29%) (Fig. 1). In total, 161 *A. marginale* strains contained one of these tandem repeats at least once and in 114 (71%) of these strains, the D20 or E20 was found in the R1 tandem repeat. Surprisingly, the highest variable amino acid was at position 20 (Fig. 4A), suggesting greater evolutionary pressure at this amino acid position. From our findings, G was the most frequent amino acid at position 20 (Fig. 4C), in both R1 and non-R1 tandem repeats (data not shown), but only 4 amino acids were found at position 20 in R1 (from highest to lowest frequency: G, D, E and serine [S]) while 7 different amino acids were found at position 20 in non-R1 tandem repeats (G, D, E, S, T, isoleucine [I] and tyrosine [Y]) (Fig. 4C). In previous experiments, non-R1 tandem repeats had a phylogenetic correlation with tick-transmissible strains, but this correlation was not seen with R1 tandem repeats [9]. We propose that non-R1 tandems are also involved in *A. marginale*-tick interactions which require more genetic variability, because more than 20 different tick species have been reported to transmit *A. marginale* [24].

Protein conformation. As proposed previously both amino acid sequence and protein conformation may contribute to the function of MSP1a as adhesin [10]. Herein, we explored this hypothesis by predicting the 2-D structure of all the MSP1a tandem repeats. We found that 14 models explained all of the variability of 2-D structure among the 193 tandem repeats (Fig. 5). Three α-helical 2-D structure models, differing in the length and amount of α helices in the tandem repeat, described 68% of the 2-D structure variation (presented as A, σ and F in Fig. 5). The

analysis revealed that the amino acid at position 20 correlated with specific 2-D structure changes in the tandem repeat. When D or E amino acids were at this position, the structure of the tandem repeat was predominantly long α-helical structures (Model types 39, A, 13 and σ), but when a G was in this position, the repeat was a short α helix, β-strand or coiled 2-D structure (Model types 4, 10, α and 48) (Fig. 5). The other four amino acids that were found at lower frequencies at position 20, (I, Y, T and S; Fig. 4C), except for Y, retained the α-helical 2-D structure (Fig. 5).

Our results suggest that the MSP1a tandem repeat 2-D structure also correlated with tick transmissibility (Table 2). Strains reported previously that were not transmitted by *Dermacentor sp.* had a predominant pattern for 2-D structure of tandem repeats of β strand, short α-helix or coiled structures, regardless of whether or not they had TCE-binding tandem repeats (Table 2). In contrast, abundant α-helices were found in tandem repeats of strains transmitted by ticks (Table 2). In the last case, as shown for the USA/Florida/G - (A, B[7]) strain, the presence of all seven TCE-binding tandem repeats did not correlate with tick-transmissibility; this Florida isolate was clearly shown to be non-infective for ticks or cultured tick cells (Table 2). However, the 2-D structure appeared to be a determinant for the biological transmission of *A. marginale*, because the Israel/Israel tailed/F - (1, F, M, 3) strains, while not having TCE-binding repeats but did have α-helices as 2-D structure, were tick transmissible (Table 2). As listed in Table 2, the data collected thus far regarding *A. marginale* transmissibility by ticks is related to the major vector *Dermacentor sp.* The complexity of the relationship between the 2-D
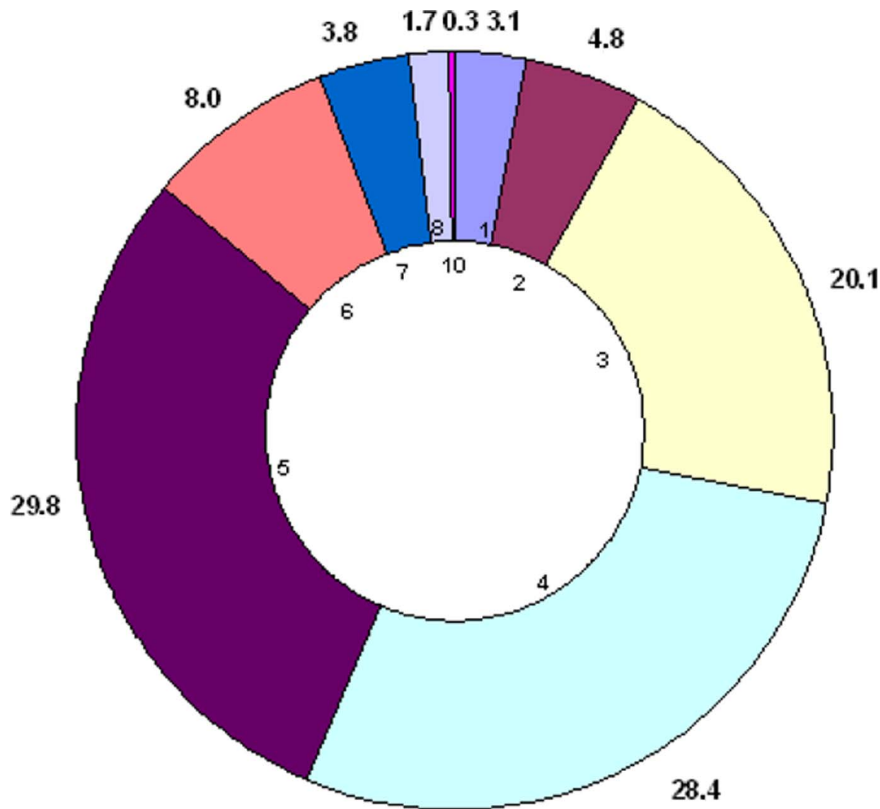
Figure 3. Number of tandem repeats among A. marginale strains. The total number of strains classified in our study were organized by the number of MSP1a tandem repeats. The percent of A. marginale strains (external numbers) containing different number of tandem repeats (internal numbers) is shown. The most common numbers of MSP1a tandem repeats among strains were 3 (yellow), 4 (light blue) and 5 (violet).
doi:10.1371/journal.pone.0065243.g003

structure, TCE-binding repeats and tick transmissibility was also seen with the Brazil/Minas Gerais/E strain–(13, 42, 13, 18) which does not contain β strands and is not transmissible by *Rhipicephalus (Boophilus) microplus* [13]. This example demonstrated a different pattern as that observed with *A. marginale* that are not transmissible by *Dermacentor sp*. The 2-D structure data presented in the present study is in agreement with an analysis performed recently on *A. marginale* MSP2 variants in tick or mammalian cells [41]. The 2-D structure analysis using PSIPRED demonstrated that MSP2 variants expressed in ticks were predominantly α-helices, while β-strands were present in MSP2 variants expressed only in mammalian cells [41,42].

Pathogen-environmental relationships. *A. marginale* was recorded in four eco-region clusters defined in our study (Table 3).

Eco-region Cluster 1 extended over large areas of central Africa and central South America, primarily Argentina and southern Brazil, and was a region with medium to high Normalized Difference Vegetation Index (NDVI) values and a well-defined seasonal decrease between June and September. The highest recorded temperature and annual rainfall of approximately 1,000 mm occurs in Eco-region Cluster 1. Eco-region Cluster 2 included vast areas of the Mesoamerican corridor, northern South America and a small territory of eastern South Africa, and included zones with high NDVI throughout the year without seasonal variability. The temperature values in Eco-region Cluster 2 were similar to those in Eco-region Cluster 1, but with an annual rainfall of approximately 1,500 mm. Eco-region Cluster 3 extended over central South Africa and scattered parts of the

Table 1. Geographical occurrence of the most common *A. marginale* strains.

Strains	Sructure of MSP1a tandem repeats	Number of strains	World occurrence
Most common	τ 22 13 18	7	4x Argentina, 3x Mexico
	α **β** β β Γ	7	4x Argentina, 2x Mexico, 1x Taiwan
Second common	34 13 4 37	6	6x South Africa
Third common	**B** B M	5	5x Argentina
	F **M** M	5	4x Argentina, 1x Mexico

The most frequent A. marginale strains and their geographical occurrence are shown. The most common tandem repeats found among all the A. marginale strains are underlined and there were found more than 60 (**M**), 80 (**β**) and 90 (**B**) times.
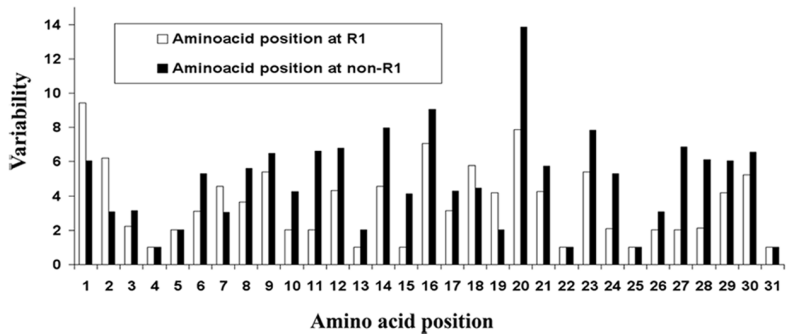doi:10.1371/journal.pone.0065243.t001

A

Amino acid variability in
Anaplasma maginale tandem repeats
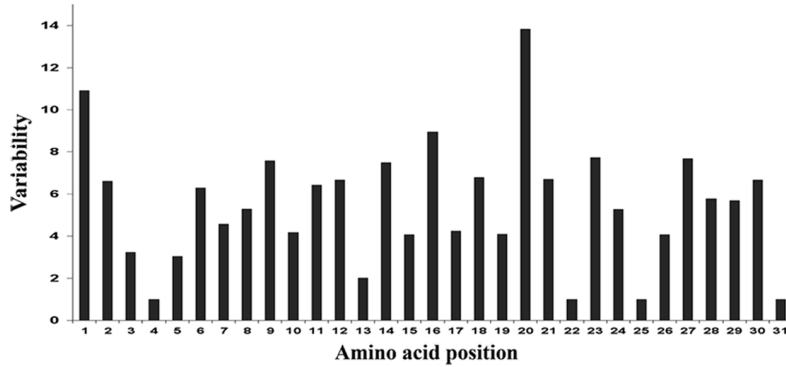
B

Comparison of amino acid variability
between R1 and non-R1 tandem repeats

C

Amino acid position	Frequency of amino acid per position																		
	W	P	V	Q	K	C	Y	F	I	R	A	T	D	L	S	N	G	H	E
1											0,54	0,39	0,04	0,005	0,005	0,005			
2												0,9			0,005	0,005	0,07	0,005	0,005
3						0,005									0,92		0,06		
4															1				
5	0,005													0,005	0,98				
6		0,01	0,005							0,95	0,01				0,005				0,01
7												0,005			0,65		0,33		
8											0,005	0,21				0,02	0,75		
9				0,92					0,02				0,005		0,005	0,01	0,02	0,02	0,01
10	0,005			0,96					0,01									0,02	
11			0,005	0,93	0,005	0,005			0,03									0,01	
12			0,005	0,01	0,01											0,005	0,05		0,89
13															0,99				0,005
14		0,005						0,023				0,029			0,8	0,017	0,12		
15			0,98							0,005						0,005			0,005
16		0,01	0,01							0,01				0,4	0,56				
17		0,04					0,005			0,005					0,94				
18		0,06		0,88	0,01				0,01						0,005			0,02	
19						0,005						0,01			0,98	0,005			
20							0,005	0,005				0,005	0,3		0,02		0,5		0,14
21		0,016		0,89						0,008			0,04					0,008	0,03
22											1								
23						0,011				0,011	0,02	0,011			0,9	0,005	0,02		
24		0,02								0,005		0,94			0,01		0,005		
25															1				
26	0,005	0,005									0,005				0,98				
27		0,017	0,005	0,91	0,005		0,005								0,017			0,03	
28			0,01				0,005			0,005				0,86	0,11				
29	0,029									0,052				0,034	0,005		0,89		
30			0,16									0,2			0,04		0,6		
31																	1		

Figure 4. Amino acid variability and frequency in A. marginale MSP1a tandem repeats. The amino acid variability (A), comparison of the variability between tandem repeats at positions R1 and non-R1 (B) and frequency (C) were calculated per amino acid position in the MSP1a tandem repeats using the formula Variability = number of different amino acids at a given position/frequency of the most common amino acid at that position [50]. The one letter amino acid code was used to name the amino acids in (C) and the most frequent amino acids per position are colored in gray.

southern USA and Mexico, and had the lowest NDVI values with minimal change across the year. This eco-region had lower temperature values and minimum rainfall. Finally, Eco-region Cluster 4 extended over large areas of the USA and had a clear NDVI signature that was low between November and March and then rose to maximum levels in July. This area was the coldest among the four eco-region clusters, with an annual rainfall of approximately 800 mm/year. The results of this study demonstrated that 82% of MSP1a R1 unique sequences were associated with only one eco-region cluster (Table 3). Seventeen R1 unique sequences (27% of the total number of R1 sequences) were reported exclusively in Eco-region Cluster 1 and shared 16 out of 31 amino acids (51.6% of the total number of amino acids) (Table 3). Sixteen R1 unique sequences (17%) were reported only in Eco-region Cluster 2 which had 64.5% identical amino acids (Table 3). Twenty-five R1 unique sequences (32%) were only found in Eco-region Cluster 3, of which 64.5% of their amino acids were shared (Table 3). Only five R1 sequences were exclusively associated with Eco-region Cluster 4, which had 77.4% identical amino acids (Table 3). Eight R1 sequences, were found simultaneously in more than one of the eco-region clusters (Table 3). These results confirmed that *A. marginale* MSP1a R1 sequences clustered according to a pattern of abiotic (climate) factors, and are related to both the species of tick vector and the performance of this tick vector in the eco-region [25]. Higher variability in R1 repeat sequences appeared in areas where several tick species are candidate vectors (i.e. USA and Canada) or where mechanical transmission is common (i.e. central Argentina). Remarkably, only one *A. marginale* MSP1a genotype has been recorded in Australia (Table S1) along with a single tick vector species, *Rhipicephalus australis* [43]. As reported previously, the hypothesis of strain geographic association was rejected [25]. Mantel's test on R1 sequences was 0.82 (P<0.001) when applied to eco-region clusters using only unique sequences. The same test provided a value of 0.31 (P = 0.145) for the distances matrix based on geographical association of strains. All the *A. marginale* MSP1a R1 sequences within each eco-region cluster appeared to be under positive selection as shown by dN/dS indexes of 1.83, 1.61, 1.54 and 1.21 for Eco-region Clusters 1 to 4, respectively. Therefore, these results confirmed the hypothesis that *A. marginale* strains are associated with factors that drive the biological performance of ticks vectors in each region [25].

Influence of a combination of factors. A phylogenetic correlation was found among *A. marginale* strains between MSP1a tandem repeats 2-D structure, transmissibility by ticks and the presence of TCE-binding tandem repeats (Fig. 6). Notably, cluster β contains all non-tick-transmissible *A. marginale* strains, abundant β-strand tandem repeat 2-D structure, and a low proportion of TCE-binding repeats (Fig. 6). The exception to this rule is the USA/St. Maries/G – (J, B^2) strain, which is tick-transmissible [34,44] but falls into this cluster. This position of the USA/St. Maries/G – (J, B^2) strain in the phylogenetic tree suggests that *A. marginale* tick-transmissible strains may evolve from non-tick-transmissible strains. The cluster α-2 contains tick-transmissible strains with the highest proportion of α-helices and all TCE-binding tandem repeats. In contrast, strains in cluster β-α-c have a more variable 2-D structure and a high proportion of TC non-

binding tandem repeats. The high β-strand content and short α-helixes in MSP1a tandem repeats appears to be associated with a non-tick-transmissible phenotype, similar to the results reported recently with MSP2 sequence study [41]. However, variable 2-D structures such as those in cluster β-α-c may be required in order to bypass the absence of TCE-binding tandem repeats and maintain the tick-transmission phenotype. The presence of TCE-binding tandem repeats could contribute to the organization of the MSP1a molecule, as seen in cluster α-1, where high content of α-helices correlated only with the presence of TCE-binding tandem repeats. Additionally, the analysis using the GeneSilico Metaserver predicted that tandem repeats have a protein disorder across the whole tandem repeat (data not shown). Intrinsically disordered proteins demonstrated better molecular recognition due to a higher specificity, larger interacting surfaces and different folding patterns upon binding [45].

Analysis of B Cell Epitope in MSP1a Tandem Repeats

Variation in *A. marginale* outer membrane proteins, such as MSP1a, is a major challenge in developing vaccines that can provide cross-protection between the diversity of strains worldwide. MSP1a has long been investigated as a vaccine candidate [68,32–34] due to the presence of a conserved neutralization-sensitive B-cell epitope at position 20–26 of tandem repeats [6,29]. However, a study [31] of the the antibody response to the strain USA/Oklahoma/G - (K, C, H), demonstrated that after vaccination with whole *A. marginale* or recombinant MSP1a, a different MSP1a B-cell epitope was immunodominant, SSAGGQQQESS, a linear epitope at amino acid positions 4 to 14 of the tandem repeat. As the antibody response is of principal importance in anaplasmosis, strain to strain variation in tandem repeat B-cell epitopes would be an important consideration in development of an MSP1a recombinant vaccine [46–48]. We therefore characterized the diversity of the immunodominant position 4–14 B-cell epitope among sequenced strains.

This epitope showed high sequence variability among all MSP1a sequences reported to date (Fig. 4A). From the 172 MSP1a tandem repeats included in the B-cell epitope analysis, 53 sequence variants were found; nevertheless 5 of those variants covered 64% of the total epitope variability (Figs. 7A and 7B). These 5 variants formed 2 phylogenetic clusters (Fig. 7C); variants in cluster 2 share the same antibody recognition site, while those in cluster 1, types 1 and 11, have different antibody recognition sites (data not shown). All B-cell epitope types were surface exposed (data not shown) as was previously predicted for the Type 1 B-cell epitope using the TMHMM2 algorithm [31].

Seven of the 53 B-cell epitope variants gave a 0 score in both B-cell epitope prediction servers BCEPRED and BCPREDS (data not shown), suggesting that some amino acid changes in the immunodominant B-cell epitope (amino acids 4–14) could be the determining factor for the loss of this epitope. Analysis by VaxiJen, a predictor of protective antigens [49], demonstrated that the highest VaxiJen score belongs to the type model B-cell epitope, while types 1, 10, 11 and 17 have VaxiJen scores lower than the type model but higher than the average for all 53 epitopes (Fig. 7D). Among the main types of B-cell epitopes, a linear but negative correlation was observed between VaxiJen and

2D Structure	Model Presented	Tandem repeats in R1 position	Tandem repeats in non-R1 position
Pred: CCCCCCCHHHHHCHHCCCCCCCCHHCCCC AA: DDSSSASGQQQESSVLSQSDQASTSSQLG	**39**	---	---
Pred: CCCCCCCCCHHHHCCCCCCCHHHHCCCC AA: DDSSSASGQQQESSVSSQSEASTSSQLG	**A**	D, E, J, L, Q, U, Σ, 5, 12, 34, 40, 41, 64, 76, 107, 136, 27	N, Γ, φ, 26, 31, 45, 47, 57, 58, 62, 66, 71, 79, 89, 97, 112, 130, 134, 135, 139
Pred: CCCCCCCHHHHHCCHHCCCCCCCCCCCCC AA: TDSSSASGQQQESSVLSQSDQASTSSQLG	**13**	103	138
Pred: CCCCCCCHHHHHHHHCCCCCCCCCCCCCC AA: ADSSSAGGQQQESIVSSQSDHASTSSQLG	σ	1, 3, 8, 23, 28, 37, 60, 67, 77, 80	14, 20, 21, 25, 29, 30, 33, 38, 44, 61, 63, 65, 68, 70, 78, 82, 83, 87, 96, 98, 104, 137
Pred: CCCCCCCHHHHCCCCCCCCCCCCCCCCCC AA: TDSSSASGQQQESSVSSQSGQASTSSQLG	**F**	G, I, M, τ, 6, 7, 19, 42, 73, 74, 84, 85, 106	H, m, P, W, β, 9, 15, 17, 18, 24, 26, 35, 43, 46, 55, 59, 72, 75, 81, 86, 90, 91, 92, 94, 95, 100, 111, 118, 119, 120, 125, 127, 131
Pred: CCCCCCCCCCCCCCCCCCCCCCHHHCCCC AA: ADSSSAGGQQHESSGSSQSDQASTSSQLW	R	T, 129	36
Pred: CCCCHHEECCCCCCCCHHHHHCCCC AA: SAGGQQQESSVSSQSDQASTSSQLG	**O**	---	---
Pred: CCCCCCCHHHHHCCEECCCCCCCCCCCCC AA: TDSSSASGQQQESSVLSQSGQASTSSQLG	**4**	53, 56, 126	110, 113, 124
Pred: CCCCCCCCHHEECCEECCCCCCCCCCCCC AA: ADSSSASGQQQESSVLSPSGQASTSSQLG	10	93	11, 22
Pred: CCCCCCCCCCEECCCCCCCCCCHHHCCCC AA: ADSSSAGGQQQESSVSSQSDQASTSSQLG	**B**	K, 88, 101, 117	V, 105
Pred: CCCCCCCCCCEECCEECCCCCCHHCCCC AA: TDSSSASGQQQVSSVSSQSDASTSSQLG	52	---	116
Pred: CCCCCCCCCCEECCEECCCCCCCCCCCCC AA: TDGSSASGQQQESSVLSPSGQASTSSQLGVG	**48**	54	---
Pred: CCCCCCCCCCEECCCCCCCCCCCCCCCCC AA: ADSSSAGGQQQESSVSSQSGQASTSSQLG	**C**	16, 121, 132	Π, 32, 99, 102, 109, 122, 123, 128, 133, 140
Pred: CCCCCCCCCCCCCCCCCCCCCCCC AA: ADSSSASGVLSQSGQASTSSQLG	**α**	69, 108	μ, 2, 49, 50, 51, 114, 115

Legend:

- ⬭ = helix
- ➡ = strand
- ─── = coil

Conf: = confidence of prediction

Pred: predicted secondary structure

AA: target sequence

Figure 5. Changes in putative 2D structure and disorder analysis of A. marginale MSP1a tandem repeats. The PSIPRED web server was used to predict the 2D structure. The tandem repeats were grouped into fourteen 2D structure models. Tandem repeats shown represent prototypes of corresponding tandem repeat 2D structures. The second column shows (model presented) the ID of the tandem repeat presented as prototype. Models ID in red represent tandem repeats in R1 position (first tandem in the MSP1a sequence).
doi:10.1371/journal.pone.0065243.g005

BCPREDS scores and between Blastp and BCPREDS scores (Figs. 7E and 7F), suggesting a relationship between sequence identity and immune properties among the B-cell epitopes. Overall, these results suggested that different immune properties exist among the different MSP1a types of the B-cell epitopes.

As this is an immunodominant epitope [31], tandem repeats with epitopes predicted to be recognised by different antibodies could be a factor in the frequent lack of cross-protection between heterologous strains. Conversely, strains which share the same type of antibody recognition site may be more likely to be cross-protective.

A correlation ($R^2 = 0.69$) was found between the number of 2-D structure models present in a given geographic location and the amount of B-cell epitope types in the same region (Fig. 2). Therefore, we explored the hypothesis that there was a link between 2-D structure and B-cell epitopes among the MSP1a tandem repeats. An α-helical structure was seen in 88% of the tandem repeats containing type 1 B-cell epitopes and in 100% of tandem repeats containing types 10, 11 or 17 B-cell epitopes. In contrast, 69% of the tandem repeats containing type model B-cell epitopes had β-strand structures. Interestingly, a correlation was found between tick transmissibility and the type of B-cell epitopes present on MSP1a repeats, possibly due to these structural differences between epitope types. 71% of the MSP1a tandem repeats present in non-tick-transmissible *A. marginale* strains were found to have type model B-cell epitopes, whereas 87% of the

tandem repeats in tick transmissible strains contained type 1 B-cell epitopes. This data suggest antigenic differences between tick-transmissible and not-transmissible *A. marginale* strains, and agrees with the finding that both type 1 and model type epitopes fall into different phylogenetic clusters (Fig. 7C) presenting different putative antibodies recognition sites. Both epitopes had the highest VaxiJen and BCPRED scores among the 5 most common B-cell epitopes, but shared low identity as shown by Blastp score (data not shown).

Collectively, the results of these studies demonstrate that the unified nomenclature proposed herein using MSP1a sequences provides information about *A. marginale* strain world distribution, transmissibility by ticks, infective potential, antigenic variability and putative utility for MSP1a vaccine development. The structural and immune analyses of MSP1a revealed a phylogenetic correlation between A. marginale tick transmissibility, 2-D structure adopted by the tandem repeats and the type of B-cell epitopes present in the tandem repeats. These results are fundamental information for design of MSP1a structure-based vaccines which would be cross protective against multiple *A. marginale* strains, and for development of serodiagnostic methods based on differential B-cell epitopes, for epidemiological characterization of field strains.

Table 2. Effect of putative MSP1a tandem repeat 2-D structure on A. marginale tick transmission phenotype.

Strains	MSP1a tandem repeats 2D structure	Transmission by ticks		
		Dermacentor spp.	R. sanguineus or R. microplus	H. excavatum
USA/Idaho/C - (D⁵, E)	(α-α, α-α, α-α, α-α, α-α, α-α)	Yes [*]	ND	ND
Puerto Rico/Puerto Rico/C - (E, φ⁵)	(α-α, α-α, α-α, α-α, α-α, α-α)	Yes [***]	Yes [***]	ND
USA/Virginia/G - (A, B)	(α-α, β-α)	Yes [*]	ND	ND
USA/St.Maries/G - (J, B²)	(α-α, β-α, β-α)	Yes [*]	Yes[***]	ND
USA/Oklahoma/G - (U)	(α-α)	Yes[+]	ND	ND
USA/Missisippi/D - (D⁴, E)	(α-α, α-α, α-α, α-α, α-α)	Yes[*]	ND	ND
USA/Rassmusen/− (A, F, H)	(α-α, α-c, α-c)	Yes[*]	ND	ND
USA/Kansas/− (E, M, φ)	(α-α, α-c, α-α)	Yes[−]	ND	ND
Nigeria/Zaria/− (54, 55, F)	(β-β, α-c, α-c)	Yes[**]	ND	ND
Israel/Israel tailed/F - (1, F, M, 3)	(α-c, α-c, α-c, α-c)	ND	Yes[****]	Yes[****]
Israel/Israel non tailed/G - (1, 4)	(α-c, α-β)	ND	Yes[****]	No[****]
USA/Florida/G - (A, B⁷)	(α-α, β-α, β-α, β-α, β-α, β-α, β-α, β-α)	No[*]	ND	ND
USA/California/G - (B², C)	(β-α, β-α, β-c)	No[*]	ND	ND
USA/Okeechobee/G - (L, B, C, B, C)	(α-α, β-α, β-c, β-α, β-c)	No[*]	ND	ND
USA/Illinois/G - (M, N, B, M, H)	(α-c, α-α, β-α, α-c, α-c)	No[*]	ND	ND

The information about transmission of A. marginale strains by ticks was collected from (*) de la Fuente et al. (2003) [10],
(**) Zivkovic et al. (2007) [65],
(***) Futse et al. (2003) [44],
(****) Shkap et al. 2009 (****) [39],
(−) Leverich et al. (2008) [66], and (+) Barbet et al (2001) [67].
TCE-binding tandem repeats are underlined. Abbreviation: ND, not determined.
doi:10.1371/journal.pone.0065243.t002

Table 3. Association of *A. marginale* MSP1a R1 sequences with world ecoregions.

Ecoregion	R1 sequences[a]	Other R1 sequences[b]
1: central Africa and central South America, primarily Argentina and southern Brazil	M, 4, 8, 12, 16, 56, 60, 64, 67, 69, 72, 78, 93, 132, γ, π, τ	A, B, D, T, 13, 23, α
2: Mesoamerican corridor, northern South America and a small territory of eastern South Africa	E, F, 28, 37, 48, 53, 54, 84, 85, 101, 117, 121, 126, 129, 136, ε	A, B, L, T, 13, 23, α
3: central South Africa and scattered parts of southern USA and Mexico	M, O, Q, U, 1, 3, 5, 6, 7, 27, 33, 34, 39, 40, 42, 74, 77, 82, 141, 142, 143, 147, 151, 154, 155,	A, D
4: USA	I, J, K, O, U, 19,	A, B, L, α

World ecoregions were built upon temporal series of NDVI values.
[a]R1 sequences recorded in one ecoregion only.
[b]R1 sequences that have been reported in other ecoregions.
doi:10.1371/journal.pone.0065243.t003

Methods

Anaplasma marginale Strains Classification

A total of 289 *A. marginale* MSP1a sequences with complete tandem repeat regions included in this study were obtained from published research and the GenBank sequence database [http://www.ncbi.nlm.nih.gov/]. These sequences were analyzed and classified, and the tandem repeats were named (or renamed) following the nomenclature proposed by Allred et al. [6] and de la Fuente et al. [14]. When microsatellite sequences were included in the *msp1a* published nucleotide sequence, they were used to assign a genotype following the system of Estrada-Peña et al. [25]. Briefly, the 5′-UTR microsatellite located between the putative Shine-Dalgarno (SD) sequence (GTAGG) and the translation initiation codon (ATG), GTAGG (**G/A TTT**) m (**GT**) n **T** ATG (microsatellite sequence is shown in bold letters) and the SD-ATG distance (d) calculated in nucleotides as $(4 \times m)+(2 \times n) +1$ were used. We propose one nomenclature for *A. marginale* strains based on MSP1a with the following structure: country/locality/microsatellite genotype - (structure of tandem repeat), and all MSP1a sequences were classified using this nomenclature. When multiple strains had 100% amino acid sequence similarity across tandem repeats, they were listed under one strain name, with geographical information taken from the isolate with the most complete information. When this information was equal between isolates, information was used from the isolate first submitted to GenBank.

Amino Acid Variability within MSP1a Tandem Repeats

Tandem repeat sequences were aligned using Clustalw, and each amino acid position was numbered from 1 to 31. The amino acid variability was determined using the formula of Kuby et al. [50]. The variability was equal to the number of different amino acids at a given position/frequency of the most common amino acid at that position.

Correlation Analysis between MSP1a Tandem Repeats and World Ecological Regions

The analysis was conducted as described previously, assuming that (i) eco-regions could be delineated by quantitative abiotic characters based on well-recognized and repeatable attributes and (ii) *A. marginale* strains were associated with each eco-region and subjected to different environmental conditions that could be analyzed by multivariate geographic clustering [25]. The feature selected to build the eco-regions was the NDVI, which is a variable that reflects vegetation stress and summarizes information about the ecological background for the performance of tick populations [25]. A 0.1° resolution series of monthly NDVI data was obtained
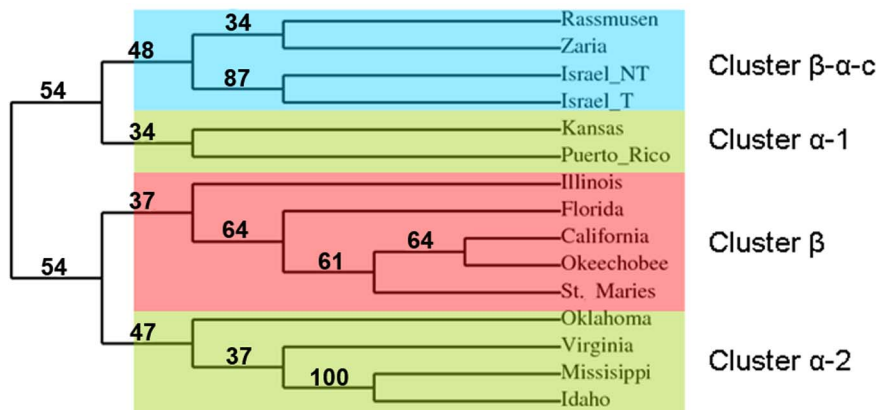
Figure 6. Phylogenetic tree based on MSP1a tandem repeat amino acid sequences. The MSP1a sequences from tick-transmissible and non-transmissible strains (Table 2) were included in the phylogenetic analysis. The phylogenetic tree was reconstructed using the neighbor joining and maximum likelihood methods. Reliability for internal branch was assessed using the bootstrapping method with 1000 bootstrap replicates. Bootstrap values are shown as % in the internal branch. The tree shows four phylogenetic clusters containing different patterns of MSP1a tandem repeat 2D structures. Cluster β-α-c (blue), cluster α-1 and cluster α-2 (beige) contain tick-transmissible *A. marginale* strains while in cluster β (red) fall the non-tick-transmissible strains.
doi:10.1371/journal.pone.0065243.g006
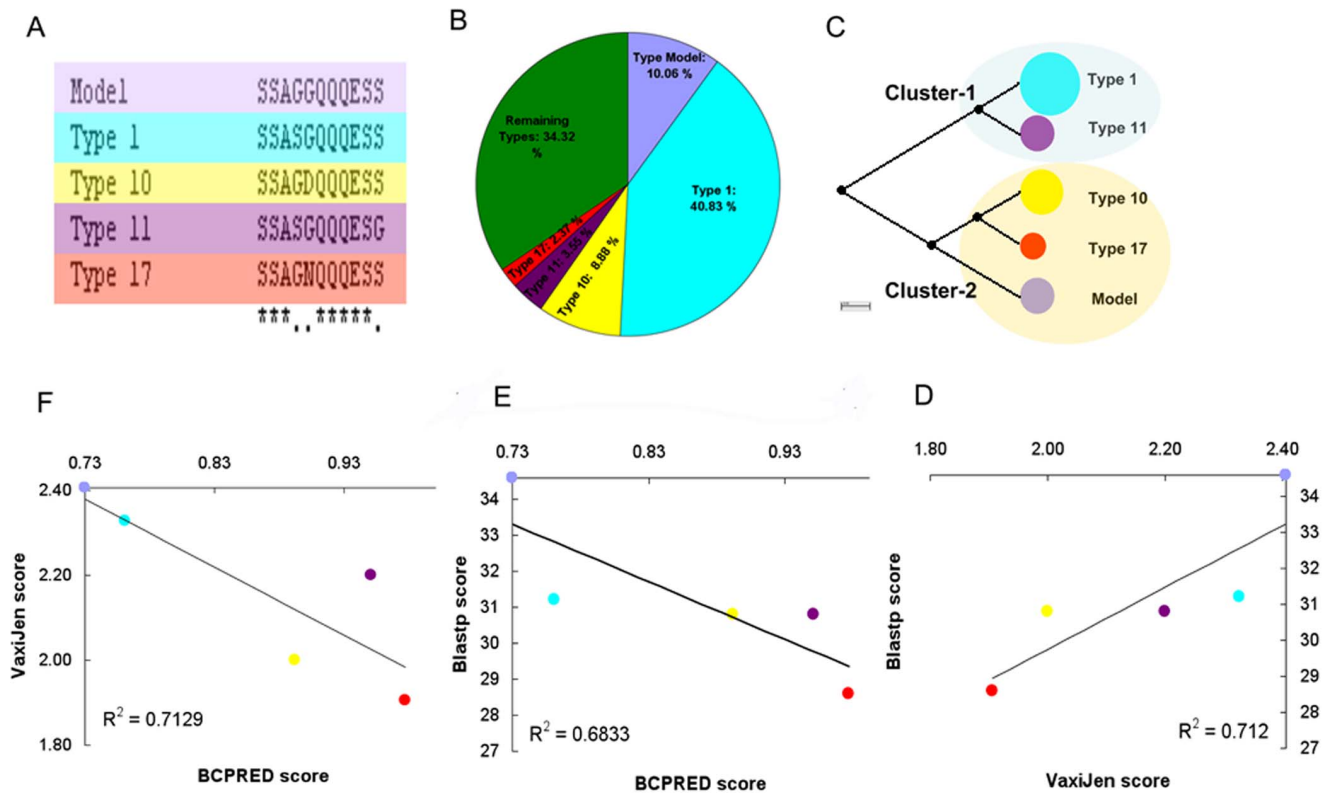
Figure 7. B-cell epitope analysis in A. marginale MSP1a tandem repeats. The B-cell epitopes were predicted using BCPRED server. The type 1 B-cell epitope was used as reference (Model) for comparisons. (A) Clustalw alignment and amino acid changes in the 5 more represented MSP1a tandem repeat B cell epitopes. B-cell epitope types model (light violet), 1 (blue), 10 (yellow), 11 (dark violet) and 17 (red) are shown. (B) Percent of tandem repeats containing each type of B cell epitopes. (C) Neighbor joining phylogenetic tree based on B cell epitope amino acid sequences showing the two clusters formed by the 5 more represented B cell epitopes. Cluster-1: Types 1 and 11 and Cluster-2: Types Model, 10 and 17. Correlations between VaxiJen/Blastp (D), BCPRED/Blastp scores (E) and VaxiJen/BCPRED (F) scores are shown. These correlations suggest that the epitopes with higher homology (Blastp score) share in common the immunogenic properties represented by VaxiJen/BCPRED.
doi:10.1371/journal.pone.0065243.g007

for the period 1986–2006. The 12 averaged monthly images were subjected to Principal Components Analysis (PCA) to obtain decomposition into the main axes representing the most significant, non-redundant information. The strongest principal axes were chosen using Cattell's Scree Test [25]. The PCA analysis retained three principal axes, including 92% of the total variance. A hierarchical agglomerative clustering on PCA values was then used to classify multiple geographical areas into a single common set of discrete regions. Mahalanobis distance was used as a measure of dissimilarity and the weighted pair-group average was used as the amalgamation method. A value of 0.05 was used as the cut-off probability for assignment to a given eco-region.

Bioinformatics

Secondary structure was predicted using the position-specific scoring matrices method [51] from the PSIPRED server [52], and protein disorder was predicted using the GeneSilico Metaserver [53].

The immunodominant B-cell epitope SSAGGQQQESS (amino acid positions 4–14), previously mapped in the *A. marginale* strain USA/Oklahoma/G - (K, C, H) MSP1a sequence [31] will be referred to as epitope "Type 1". The variability among MSP1a tandem repeats within this B-cell epitope (amino acid positions 4–14) was evaluated. The percent of amino acid identity and Blastp score among the B-cell epitopes had a linear correlation ($R^2 = 0.85$), so the Blastp score was used as an identity index in

the analysis. Prediction/score of B-cell epitope was determined using BCPREDS server [54] and the protective potential of the B-cell epitope was predicted using the VaxiJen server [55]. Prediction of physicochemical properties of the B-cell epitope was assayed using BCEPRED server [56]. PepSurf algorithm [57], implemented in the PEPITOPE server [58], was used to determine the structure/position of the affinity-selected B-cell epitopes in a model protein. The 3D analysis of MSP1a tandem repeat B-cell epitopes was performed using a model of the crystal structure of the Fv corresponding with the anti-blood group A antibody AC1001 (PDB ID: 1JV5) [59].

For phylogenetic analysis, sequences were aligned with MUS-CLE (v3.7) configured for the highest accuracy [60]. After alignment, ambiguous regions (i.e., containing gaps and/or poorly aligned) were removed with Gblocks (v0.91b) [61]. The phylogenetic tree was reconstructed using the neighbor joining (NJ) and maximum likelihood methods implemented in PHYLIP package (v3.66), NJ distances were calculated using FastDist [62,63]. Reliability for internal branch was assessed using the bootstrapping method (1000 bootstrap replicates). Graphical representation and editing of the phylogenetic tree were performed with TreeDyn (v198.3) [64].

Supporting Information

Table S1 Classification of *A. marginale* strains based on the proposed nomenclature. A total of 289 MSP1a

sequences were analyzed. A. marginale 224 unique strains were classified using the nomenclature proposed in our study: Country/Locality/microsatellite genotype - (structure of tandem repeat). The 5′UTR microsatellite genotype was included when available. The structure of tandem repeats was represented following the nomenclature previously proposed [14] (Fig. 1). When the same repeat was present more than one time, a super-index was used to represent copy number for this repeat.
(PDF)

References

1. Kocan KM, de la Fuente J, Guglielmone AA, Melendez RD (2003) Antigens and alternatives for control of Anaplasma marginale infection in cattle. Clin Microbiol Rev 16: 698–712.
2. Kocan KM, de la Fuente J, Blouin EF, Garcia-Garcia JC (2004) Anaplasma marginale (Rickettsiales: Anaplasmataceae): recent advances in defining host-pathogen adaptations of a tick-borne rickettsia. Parasitology 129 Suppl: S285–300.
3. Aubry P, Geale DW (2011) A review of bovine anaplasmosis. Transbound Emerg Dis 58: 1–30.
4. Smith RD, Levy MG, Kuhlenschmidt MS, Adams JH, Rzechula DL, et al. (1986) Isolate of Anaplasma marginale not transmitted by ticks. Am J Vet Res 47: 127–129.
5. Wickwire KB, Kocan KM, Barron SJ, Ewing SA, Smith RD, et al. (1987) Infectivity of three Anaplasma marginale isolates for Dermacentor andersoni. Am J Vet Res 48: 96–99.
6. Allred DR, McGuire TC, Palmer GH, Leib SR, Harkins TM, et al. (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.
7. Rodriguez SD, Garcia Ortiz MA, Hernandez Salgado G, Santos Cerda NA, Aboytes Torre R, et al. (2000) Anaplasma marginale inactivated vaccine: dose titration against a homologous challenge. Comp Immunol Microbiol Infect Dis 23: 239–252.
8. de la Fuente J, Garcia-Garcia JC, Blouin EF, McEwen BR, Clawson D, et al. (2001) Major surface protein 1a effects tick infection and transmission of Anaplasma marginale. Int J Parasitol 31: 1705–1714.
9. de La Fuente J, Garcia-Garcia JC, Blouin EF, Rodriguez SD, Garcia MA, et al. (2001) Evolution and function of tandem repeats in the major surface protein 1a of the ehrlichial pathogen Anaplasma marginale. Anim Health Res Rev 2: 163–173.
10. de la Fuente J, Garcia-Garcia JC, Blouin EF, Kocan KM (2003) Characterization of the functional domain of major surface protein 1a involved in adhesion of the rickettsia Anaplasma marginale to host cells. Vet Microbiol 91: 265–283.
11. de la Fuente J, Torina A, Naranjo V, Caracappa S, Vicente J, et al. (2005) Genetic diversity of Anaplasma marginale strains from cattle farms in the province of Palermo, Sicily. J Vet Med B Infect Dis Vet Public Health 52: 226–229.
12. Palmer GH, Rurangirwa FR, McElwain TF (2001) Strain composition of the ehrlichia Anaplasma marginale within persistently infected cattle, a mammalian reservoir for tick transmission. J Clin Microbiol 39: 631–635.
13. Ruiz PM, Passos LM, Ribeiro MF (2005) Lack of infectivity of a Brazilian Anaplasma marginale isolate for Boophilus microplus ticks. Vet Parasitol 128: 325–331.
14. de la Fuente J, Ruybal P, Mtshali MS, Naranjo V, Shuqing L, et al. (2007) Analysis of world strains of Anaplasma marginale using major surface protein 1a repeat sequences. Vet Microbiol 119: 382–390.
15. Barbet AF, Palmer GH, Myler PJ, McGuire TC (1987) Characterization of an immunoprotective protein complex of Anaplasma marginale by cloning and expression of the gene coding for polypeptide Am105L. Infect Immun 55: 2428–2435.
16. Palmer GH, Rurangirwa FR, Kocan KM, Brown WC (1999) Molecular basis for vaccine development against the ehrlichial pathogen Anaplasma marginale. Parasitol Today 15: 281–286.
17. Kocan KM, de la Fuente J (2003) Co-feeding studies of ticks infected with Anaplasma marginale. Vet Parasitol 112: 295–305.
18. Barbet AF, Blentlinger R, Yi J, Lundgren AM, Blouin EF, et al. (1999) Comparison of surface proteins of Anaplasma marginale grown in tick cell culture, tick salivary glands, and cattle. Infect Immun 67: 102–107.
19. Bowie MV, de la Fuente J, Kocan KM, Blouin EF, Barbet AF (2002) Conservation of major surface protein 1 genes of Anaplasma marginale during cyclic transmission between ticks and cattle. Gene 282: 95–102.
20. Ueti MW, Reagan JO, Jr., Knowles DP, Jr., Scoles GA, Shkap V, et al. (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. de La Fuente J, Passos LM, Van Den Bussche RA, Ribeiro MF, Facury-Filho EJ, et al. (2004) Genetic diversity and molecular phylogeny of Anaplasma marginale isolates from Minas Gerais, Brazil. Vet Parasitol 121: 307–316.
22. de la Fuente J, Lew A, Lutz H, Meli ML, Hofmann-Lehmann R, et al. (2005) Genetic diversity of anaplasma species major surface proteins and implications for anaplasmosis serodiagnosis and vaccine development. Anim Health Res Rev 6: 75–89.
23. de la Fuente J, Kocan KM, Blouin EF, Zivkovic Z, Naranjo V, et al. (2010) Functional genomics and evolution of tick-Anaplasma interactions and vaccine development. Vet Parasitol 167: 175–186.
24. Kocan KM, de la Fuente J, Blouin EF, Coetzee JF, Ewing SA (2010) The natural history of Anaplasma marginale. Vet Parasitol 167: 95–107.
25. Estrada-Peña A, Naranjo V, Acevedo-Whitehouse K, Mangold AJ, Kocan KM, et al. (2009) Phylogeographic analysis reveals association of tick-borne pathogen, Anaplasma marginale, MSP1a sequences with ecological traits affecting tick vector performance. BMC Biol 7: 57.
26. Brown WC, Palmer GH, Lewin HA, McGuire TC (2001) CD4(+) T lymphocytes from calves immunized with Anaplasma marginale major surface protein 1 (MSP1), a heteromeric complex of MSP1a and MSP1b, preferentially recognize the MSP1a carboxyl terminus that is conserved among strains. Infect Immun 69: 6853–6862.
27. Brown WC, McGuire TC, Zhu D, Lewin HA, Sosnow J, et al. (2001) Highly conserved regions of the immunodominant major surface protein 2 of the genogroup II ehrlichial pathogen Anaplasma marginale are rich in naturally derived CD4+ T lymphocyte epitopes that elicit strong recall responses. J Immunol 166: 1114–1124.
28. Brown WC, McGuire TC, Mwangi W, Kegerreis KA, Macmillan H, et al. (2002) Major histocompatibility complex class II DR-restricted memory CD4(+) T lymphocytes recognize conserved immunodominant epitopes of Anaplasma marginale major surface protein 1a. Infect Immun 70: 5521–5532.
29. Palmer GH, Waghela SD, Barbet AF, Davis WC, McGuire TC (1987) Characterization of a neutralization-sensitive epitope on the Am 105 surface protein of Anaplasma marginale. Int J Parasitol 17: 1279–1285.
30. Oberle SM, Palmer GH, Barbet AF, McGuire TC (1988) Molecular size variations in an immunoprotective protein complex among isolates of Anaplasma marginale. Infect Immun 56: 1567–1573.
31. Garcia-Garcia JC, de la Fuente J, Kocan KM, Blouin EF, Halbur T, et al. (2004) Mapping of B-cell epitopes in the N-terminal repeated peptides of Anaplasma marginale major surface protein 1a and characterization of the humoral immune response of cattle immunized with recombinant and whole organism antigens. Vet Immunol Immunopathol 98: 137–151.
32. Palmer GH, Oberle SM, Barbet AF, Goff WL, Davis WC, et al. (1988) Immunization of cattle with a 36-kilodalton surface protein induces protection against homologous and heterologous Anaplasma marginale challenge. Infect Immun 56: 1526–1531.
33. Palmer GH, Barbet AF, Cantor GH, McGuire TC (1989) Immunization of cattle with the MSP-1 surface protein complex induces protection against a structurally variant Anaplasma marginale isolate. Infect Immun 57: 3666–3669.
34. de la Fuente J, Kocan KM, Garcia-Garcia JC, Blouin EF, Halbur T, et al. (2003) Immunization Against Anaplasma marginale Major Surface Protein 1a Reduces Infectivity for Ticks. The International Journal of Applied Research in Veterinary Medicine 1.
35. Blouin EF, Saliki JT, de la Fuente J, Garcia-Garcia JC, Kocan KM (2003) Antibodies to Anaplasma marginale major surface proteins 1a and 1b inhibit infectivity for cultured tick cells. Vet Parasitol 111: 247–260.
36. McGarey DJ, Barbet AF, Palmer GH, McGuire TC, Allred DR (1994) Putative adhesins of Anaplasma marginale: major surface polypeptides 1a and 1b. Infect Immun 62: 4594–4601.
37. Lew AE, Bock RE, Minchin CM, Masaka S (2002) A msp1alpha polymerase chain reaction assay for specific detection and differentiation of Anaplasma marginale isolates. Vet Microbiol 86: 325–335.
38. de la Fuente J, Garcia-Garcia JC, Blouin EF, Kocan KM (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.
39. Shkap V, Kocan K, Molad T, Mazuz M, Leibovich B, et al. (2009) Experimental transmission of field Anaplasma marginale and the A. centrale vaccine strain by Hyalomma excavatum, Rhipicephalus sanguineus and Rhipicephalus (Boophilus) annulatus ticks. Vet Microbiol 134: 254–260.
40. Benz I, Schmidt MA (2002) Never say never again: protein glycosylation in pathogenic bacteria. Mol Microbiol 45: 267–276.
41. Chavez AS, Felsheim RF, Kurtti TJ, Ku PS, Brayton KA, et al. (2012) Expression patterns of Anaplasma marginale Msp2 variants change in response to growth in cattle, and tick cells versus mammalian cells. PLoS One 7: e36012.

Author Contributions

Conceived and designed the experiments: JdlF AC-C JJV LMFP LG. Performed the experiments: JJV RK KL JF MT AEP AC-C. Analyzed the data: JJV RK KL JF MT AE-P AC-C. Contributed reagents/materials/analysis tools: EZ VS MR. Wrote the paper: JdlF KMK JF RK KL JJV AE-P MT AC-C.

42. Futse JE, Brayton KA, Nydam SD, Palmer GH (2009) Generation of antigenic variants via gene conversion: Evidence for recombination fitness selection at the locus level in *Anaplasma marginale*. Infect Immun 77: 3181–3187.

43. Estrada-Peña A, Venzal JM, Nava S, Mangold A, Guglielmone AA, et al. (2012) Reinstatement of *Rhipicephalus (Boophilus) australis* (Acari: Ixodidae) with redescription of the adult and larval stages. J Med Entomol 49: 794–802.

44. Futse JE, Ueti MW, Knowles DP, Jr., Palmer GH (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.

45. Dunker AK, Brown CJ, Lawson JD, Iakoucheva LM, Obradovic Z (2002) Intrinsic disorder and protein function. Biochemistry 41: 6573–6582.

46. Valdez RA, McGuire TC, Brown WC, Davis WC, Jordan JM, et al. (2002) Selective in vivo depletion of CD4(+) T lymphocytes with anti-CD4 monoclonal antibody during acute infection of calves with *Anaplasma marginale*. Clin Diagn Lab Immunol 9: 417–424.

47. Santos PS, Nascimento R, Rodrigues LP, Santos FA, Faria PC, et al. (2012) Functional epitope core motif of the *Anaplasma marginale* major surface protein 1a and its incorporation onto bioelectrodes for antibody detection. PLoS One 7: e33045.

48. Suarez CE, Noh S (2011) Emerging perspectives in the research of bovine babesiosis and anaplasmosis. Vet Parasitol 180: 109–125.

49. Maritz-Olivier C, van Zyl W, Stutzer C (2012) A systematic, functional genomics, and reverse vaccinology approach to the identification of vaccine candidates in the cattle tick, *Rhipicephalus microplus*. Ticks Tick Borne Dis 3: 179–187.

50. Kindt TJ, Goldsby RA, Osborne BA, Kuby J (2007) Kuby immunology. New York: W.H. Freeman. xxii, 574, A-531, G-512, AN-527, I-527 p.

51. Jones DT (1999) Protein secondary structure prediction based on position-specific scoring matrices. J Mol Biol 292: 195–202.

52. Buchan DW, Ward SM, Lobley AE, Nugent TC, Bryson K, et al. (2010) Protein annotation and modelling servers at University College London. Nucleic Acids Res 38: W563–568.

53. Kurowski MA, Bujnicki JM (2003) GeneSilico protein structure prediction meta-server. Nucleic Acids Res 31: 3305–3307.

54. El-Manzalawy Y, Dobbs D, Honavar V (2008) Predicting linear B-cell epitopes using string kernels. J Mol Recognit 21: 243–255.

55. Doytchinova IA, Flower DR (2007) VaxiJen: a server for prediction of protective antigens, tumour antigens and subunit vaccines. BMC Bioinformatics 8: 4.

56. Saha S, Raghava GP (2006) Prediction of continuous B-cell epitopes in an antigen using recurrent neural network. Proteins 65: 40–48.

57. Mayrose I, Shlomi T, Rubinstein ND, Gershoni JM, Ruppin E, et al. (2007) Epitope mapping using combinatorial phage-display libraries: a graph-based algorithm. Nucleic Acids Res 35: 69–78.

58. Mayrose I, Penn O, Erez E, Rubinstein ND, Shlomi T, et al. (2007) Pepitope: epitope mapping from affinity-selected peptides. Bioinformatics 23: 3244–3246.

59. Thomas R, Patenaude SI, MacKenzie CR, To R, Hirama T, et al. (2002) Structure of an anti-blood group A Fv and improvement of its binding affinity without loss of specificity. J Biol Chem 277: 2059–2064.

60. Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792–1797.

61. Castresana J (2000) Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol Biol Evol 17: 540–552.

62. Elias I, Lagergren J (2007) Fast computation of distance estimators. BMC Bioinformatics 8: 89.

63. Felsenstein J (1989) Mathematics vs. Evolution: Mathematical Evolutionary Theory. Science 246: 941–942.

64. Chevenet F, Brun C, Banuls AL, Jacq B, Christen R (2006) TreeDyn: towards dynamic graphics and annotations for analyses of trees. BMC Bioinformatics 7: 439.

65. Zivkovic Z, Nijhof AM, de la Fuente J, Kocan KM, Jongejan F (2007) Experimental transmission of *Anaplasma marginale* by male *Dermacentor reticulatus*. BMC Vet Res 3: 32.

66. Leverich CK, Palmer GH, Knowles DP, Jr., Brayton KA (2008) Tick-borne transmission of two genetically distinct *Anaplasma marginale* strains following superinfection of the mammalian reservoir host. Infect Immun 76: 4066–4070.

67. Barbet AF, Yi J, Lundgren A, McEwen BR, Blouin EF, et al. (2001) Antigenic variation of *Anaplasma marginale:* major surface protein 2 diversity during cyclic transmission between ticks and cattle. Infect Immun 69: 3057–3066.

68. Palmer GH, Barbet AF, Davis WC, McGuire TC (1986) Immunization with an isolate-common surface protein protects cattle against anaplasmosis. Science 231: 1299–1302.