

Article

Not peer-reviewed version

Generation of Population-Level Diversity in *Anaplasma phagocytophilum* *msp2/p44* Gene Repertoires Through Recombination

[Anthony F Barbet](#) , [David R Allred](#) ^{*} , Francy L Crosby

Posted Date: 11 February 2025

doi: 10.20944/preprints202502.0781.v1

Keywords: Tick-borne diseases; antigenic variation; repertoire variability; anaplasmosis; emerging diseases; *msp2/p44*



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC.

Copyright: This open access article is published under a Creative Commons CC BY 4.0 license, which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Article

Generation of Population-Level Diversity in *Anaplasma phagocytophilum* *msp2/p44* Gene Repertoires Through Recombination

Anthony F. Barbet ^{1,2,3,§}, David R. Allred ^{1,2,3,†,*} and Francy L. Crosby ^{1,†,¶}

¹ Department of Infectious Diseases and Immunology, University of Florida, Gainesville, FL

² Emerging Pathogens Institute, University of Florida, Gainesville, FL

³ Genetics Institute, University of Florida, Gainesville, FL

* Correspondence: allredd@ufl.edu (dra); Tel.: 1-352-294-4126 (dra)

† These authors contributed equally.

§ Deceased.

¶ Current address: Department of Veterinary Pathobiology, University of Missouri, Columbia, MO.

Abstract: *Anaplasma phagocytophilum*, a tick-borne Rickettsiales, causes the emerging disease of humans and animals, granulocytic anaplasmosis. The organism expresses an immunodominant surface protein, MSP2/P44, that undergoes rapid antigenic variation during single infections due to gene conversion of a single genomic expression site with sequences from one of ~100 transcriptionally silent genes known as “functional pseudogenes”. Most studies have indicated that the predominant gene conversion mechanism is insertion of complete central variable regions (CVRs) into the *msp2/p44* expression site via homologous recombination through 5′ and 3′ conserved regions. This suggests that it is possible that persistent infections by one strain may be self-limiting due to exhaustion of the antigenic repertoire. However, if there is substantial recombination within the functional pseudogene repertoires themselves it is likely that these repertoires may have a high rate of change. This was investigated here by analyzing the repertoires of *msp2/p44* functional pseudogenes in genome-sequenced *A. phagocytophilum* from widely different geographic locations in the USA and Europe. The data support a high probability of recombination events having occurred within and between *msp2/p44* repertoires that is not limited to the 5′ and 3′ conserved regions of the CVR, greatly expanding total potential variation. Continual variation of *msp2/p44* repertoires is predicted to aid the organism in overcoming existing immunity in the individual and causing superinfections among immune populations, and may facilitate adaptation of the microorganism to infect and cause disease in different species.

Keywords: Tick-borne diseases; antigenic variation; repertoire variability; anaplasmosis; emerging diseases; *msp2/p44*

1. Introduction

Anaplasma phagocytophilum is a Rickettsiales organism that causes an emerging, tick-borne, disease of humans characterized by >5000 cases/year in the USA (<https://www.cdc.gov/anaplasmosis/stats/index.html>). It also infects rodents, ruminants, dogs and horses worldwide and has long been a barrier to domestic sheep production in Norway (1). The disease can be persistent, a characteristic thought to be enabled in part by the sequential expression of variable surface antigens encoded by the *msp2/p44* multigene family (2). As in many systems of antigenic variation, antibody to *msp2/p44* neutralizes infection caused by the homologous serotype. However, structurally different MSP2/P44 proteins are created and dominate in different organism peaks during infection that are only recognized by antibodies generated subsequent to their appearance (3-6). The demonstrated basis for antigenic variation is insertion of a central variable

region (CVR) into an *msp2/p44* genomic expression site by gene conversion, utilizing the RecF recombination pathway (7-10). Different CVR's are present in numerous copies in the *A. phagocytophilum* genome, flanked by conserved 5' and 3' sequences. Genome sequencing of the HZ strain identified 113 copies of *msp2/p44*, some of which did not contain either or both of the 5' and 3' conserved regions (11). In *Anaplasma marginale*, which has a similar system for variation of *msp2*, there are only 7 or 8 copies of the *msp2* CVR (termed "functional pseudogenes") available for insertion into the single expression site (12, 13). In that species additional variation is achieved by the use of different short regions from the CVR's to form complex mosaics, particularly in long-term persistent infections (14, 15). In serial infections with *A. phagocytophilum* in a mouse model, among 263 expressed pseudogenes only 3 mosaics were detected and these involved contributions from only 2 different pseudogenes in each case (16). This agrees with previous data suggesting that the primary mechanism of gene conversion was insertion of a complete CVR into the expression site (17). This was confirmed using cloned *A. phagocytophilum* to infect horses and SCID mice that showed recombination break points only in the conserved 5' and 3' regions of *msp2/p44* copies and suggested that the *msp2/p44* antigenic repertoire was limited (8). This has led to the idea that long-term infections with *A. phagocytophilum* could be self-limiting, due to exhaustion of the *msp2/p44* repertoire, unless there is accompanying variation of the repertoire itself (16, 17). The availability of 28 genome-sequenced strains of *A. phagocytophilum* from different geographic locations (18) presents an opportunity to examine this possibility at the population level. Specifically, is there evidence for recombination among members of the *msp2/p44* repertoire leading to generation of diversity in the repertoires themselves? We provide here evidence of such recombination that may help to explain superinfections and persistence of this microorganism, and perhaps adaptation to novel hosts.

2. Materials and Methods

2.1. Determination of *msp2/p44* Repertoires

The repertoires of *msp2/p44* genes in each genome-sequenced strain were determined as described (18). Briefly, we used an 11 nucleotide sequence present in the 5' conserved sequence of *msp2/p44* to extract all instances of this sequence plus the downstream 469 nucleotides from all *A. phagocytophilum* genomes. A filter was then applied to verify that each gene encoded at least one of these known protein characteristics: N-terminal KELAY and N- or C-terminal LAKT amino acid motifs. The 113 *msp2/p44* gene loci previously described in the HZ strain (11) includes genes characterized as either: full-length, silent/reserved, truncated, or fragments. The above methods detected 83 *msp2/p44* genes in our re-sequenced HZ2 strain (accession #CP006616; designated HZ2_NY herein) and would not detect partial genes with no 5' or 3' conserved region, thought to be necessary for recombination into the MSP2 expression site. These selection criteria were similarly applied to *msp2/p44* genes from the human-derived Web_WI strain (accession #LANS00000000; designated ApWebster_WI) (a total of 166 genes), two horse-derived derived strains, Horse1_CA (accession #FLMF00000000) and Horse1_MN (accession #FLMC00000000) (166 genes), and two Norwegian sheep-derived strains, NorShV1 (accession #CP046639) and ApSheep_Norv2 (accession #CP015376) (172 genes). In total, 504 *msp2/p44* gene sequences were available for analysis. All sequences are provided in Supplementary Table S1.

2.2. Detection of Recombination

Sequences were aligned with CLC Bio proprietary multiple sequence alignment module (break cost = 10, cost to extend = 1) for ease of alignment editing. Alignments were manually optimized prior to use in analysis of recombination. Recombination detection was performed on alignments, using RDP5 software (19). For consistency with the demonstrated mechanism of gene conversion (7-10) the GENECONV module implemented within RDP5 was employed for analytical screening of all samples, providing multiple comparisons of linear sequences with Bonferroni correction and a highest acceptable p-value of 0.05. Individual samples were further analyzed with the integrated

modules RDP (31), Bootscan (32), Maxchi (33), Chimaera (34), SiScan (35), PhylPro (36), LARD (37), and 3Seq (38) for identification of recombination events and/or breakpoint sites, as implemented within RDP5 (19). Modules were adjusted for sensitivity at observed nucleotide change rates. The output of all detected recombinants and the statistical support for them is provided in an Excel-compatible file in Supplementary Table S2. Breakpoint density plots utilized a sequence window size of 100 nucleotides and 1000 permutations to infer the existence of statistically supported recombination hot- and cold-spots. These are presented herein as breakpoint p-density plots of probabilities in which 99% (dark grey) and 95% (light grey) confidence intervals are also shown as shaded areas. Hot-spots for recombination are inferred where the black plot lines emerge above the shaded areas and corresponding areas of low recombination (cold-spots) are suggested by plot lines dropping below the shaded areas.

2.3. Polypeptide Structural Comparisons

msp2/p44 repertoire sequences were translated into predicted MSP2 polypeptides in reading-frame 1. Sequences were submitted to the Robetta server (<http://robetta.bakerlab.org/>) for structural prediction, using the Robetta algorithm (21). Structural predictions were saved as .pdb files and visualized with UCSF Chimera software v. 1.14 (39). Superimpositions of predicted structures was accomplished with the Matchmaker module of Chimera.

3. Results

A prior study identified the individual *msp2/p44* genes comprising the repertoires of 28 genome-sequenced strains of *A. phagocytophilum* (Supplementary Table S2 in ref. (18)). Those genes sharing at least 99% identity at the nucleotide level were identified, enabling comparisons of the overall variability in *msp2/p44* repertoires present in each strain. Some strains had nearly identical repertoires, whereas other strain repertoires were completely different, and these differences were based partly on the geographic origin of each strain. For example, two strains isolated from humans in New York state shared most of their repertoires whereas those two strains shared only ~50% of their repertoires with human-derived strains from the Midwest USA. A horse-derived strain from Minnesota shared only ~1% of its repertoire with one from California. In the present study, the *msp2/p44* repertoires (Supplementary Table S1) were analyzed with RDP5 (Recombination Detection Program) software (19) to determine whether there is evidence of recombination between different repertoire genes.

Alignment of the gene repertoires from human-, horse-, or sheep-derived strains of *A. phagocytophilum* identified similar conserved and variable regions in each case (Supplementary Figure S1), suggesting that the overall structures of the genes comprising these repertoires were consistent and maintained across strains. The conserved and variable regions also conformed to what has been observed previously in different expressed *msp2/p44* cDNAs found in human patients infected with *A. phagocytophilum* (10). Indeed, in alignments of *msp2/p44* genes of all strains the conservation of structure and flanking sequences is clear (Supplementary Figure S2). The longest conserved regions were in 5' and 3' flanking regions of the genes that have been identified previously as preferred sites for recombination into the *msp2/p44* expression site as a part of antigenic variation (8). Comparing the repertoires of two human-derived strains from either New York state or Wisconsin, which shared 54% of their repertoires, confirmed the 5' flanking region as a hot-spot for recombination (Figure 1A). Although the 3' flanking region was not an obvious recombination hot-spot in this analysis, examples of recombination were observed there (e.g., Figure 1A, HZ2_NY2014 alignment). Moreover, in an analysis of all *msp2/p44* genes included in this study (Supplementary File 1, Supplementary Figure S2) very strong statistical support for the 3' recombination hot-spot was obtained (Figure 2). The recombination detected was both between individual genes present in the same repertoire and between genes present in either the New York or Wisconsin strains. Interestingly, the same gene in the HZ2_NY repertoire (1399) appeared to have contributed segments to at least two different copies (2026 and 2059) in the Web_WI repertoire. Putative recombinants also

extended beyond the 3' conserved flanking region into the 3' variable region (e.g., recombinant HZ2_NY1391; Supplementary Table S2). A similar result was obtained comparing the *msp2/p44* gene repertoires found in the two horse-derived strains of *A. phagocytophilum* from either Minnesota or California, although in this case the clear presence of a 3' recombination hot-spot was strongly supported statistically (Figure 1B). In the two sheep-derived strains from different regions of Norway the putative recombination events appeared to be more complex and extended further into 3' variable regions of the repertoires. Similar to the human isolates the 3' recombination hot-spot was not obvious, perhaps because of the resolution of recombination intermediates over a longer region (Figure 1C). In all strains the sequences showed evidence of prior recombination with unknown *msp2/p44* gene forms that were not recovered among the specific genomes sequenced, providing evidence of additional undefined diversity among *A. phagocytophilum* strains circulating in the environment. Significantly, it was possible to detect high probability recombination events between all human, horse, and sheep strains, in all combinations (Supplementary Table S2). Interestingly, high probability recombination events were detected between Norwegian sheep strain genes and those of the HZ2_NY strain, and in this case involved the 3' conserved sequences (Figure 3). The finding of recombination events among geographically broadly-distributed strains indicates there is conservation of variable sequence elements in the silent repertoire during the geographic distribution of this agent, as well as their recombination to broaden diversity. In all scenarios, the 5'-conserved flanking region was observed to be a hot-spot for recombination, and a cold-spot with low probability for recombination was maintained immediately 3' to the 5' hot-spot. The relative inconsistency of the 3' conserved flanking region as an obvious recombination hot-spot is curious, and seems to be associated with the host species from which the *A. phagocytophilum* strains were isolated. This may reflect a greater or lesser importance of sequences in that region of the MSP2 protein for interactions with specific hosts, resulting in differences in the levels of immune selection and retention of recombinants altered in that region.

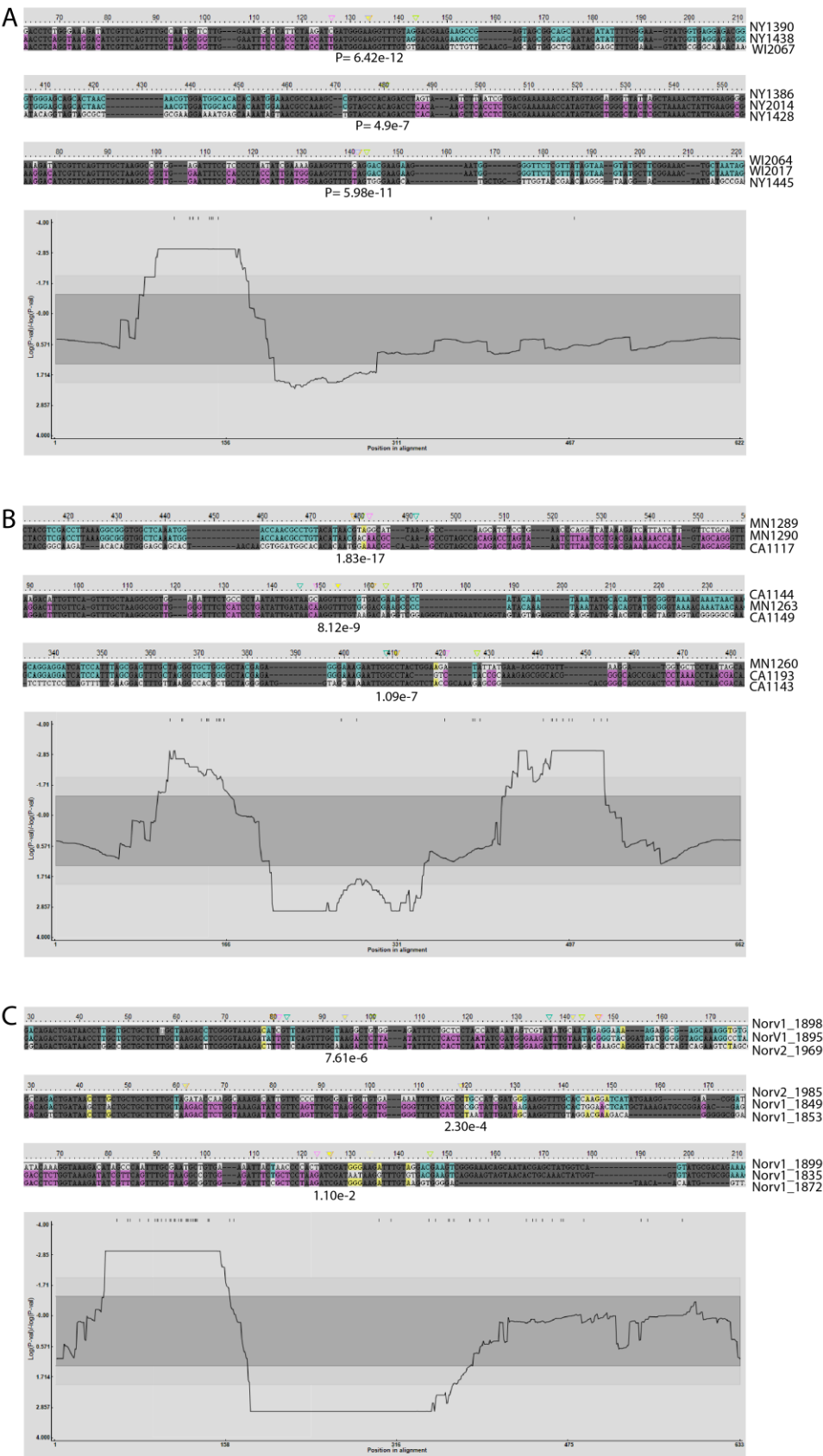


Figure 1. Predicted intra- and inter-strain recombination events between *msp2/p44* genes of *A. phagocytophilum* strains, based upon host.Evidence of recombination is provided for (A) human-derived strains, HZ2_NY and Webster_WI, (B) horse-derived strains, Horse1_CA and Horse1_MN, and (C) sheep-derived strains, NorShV1 and NorShV2. Three examples of areas of recombination between genes of strains infecting the same host and predicted with high probability are shown in each case. The top sequence of each alignment was predicted to be

the “major parent” (contributing most of the sequence), the bottom to be the “minor parent” (contributing less), and the center sequence is the predicted recombinant gene. Positions of recombination sites used in the analysis are shown above the plot, with predicted breakpoint sites indicated by inverted triangles above each alignment (yellow triangles indicate sites predicted by GENECONV). Probability values are derived from the GENECONV analysis and indicated beneath the predicted breakpoint site. A breakpoint P-density plot is provided for each analysis in which the plotted values for the alignment of all human, horse, or sheep isolate-derived sequences correspond to probabilities that recombination breakpoints are not significantly clustered. The central shaded areas indicate the 95% and 99% confidence intervals for the expected degrees of breakpoint clustering in the absence of recombination hot- and cold-spots. The fasta sequences used in the alignments are provided in Supplementary Table S1. Statistical results for the full series of recombination and breakpoint analyses performed on all alignments for all combinations are presented in Supplementary Table S2.

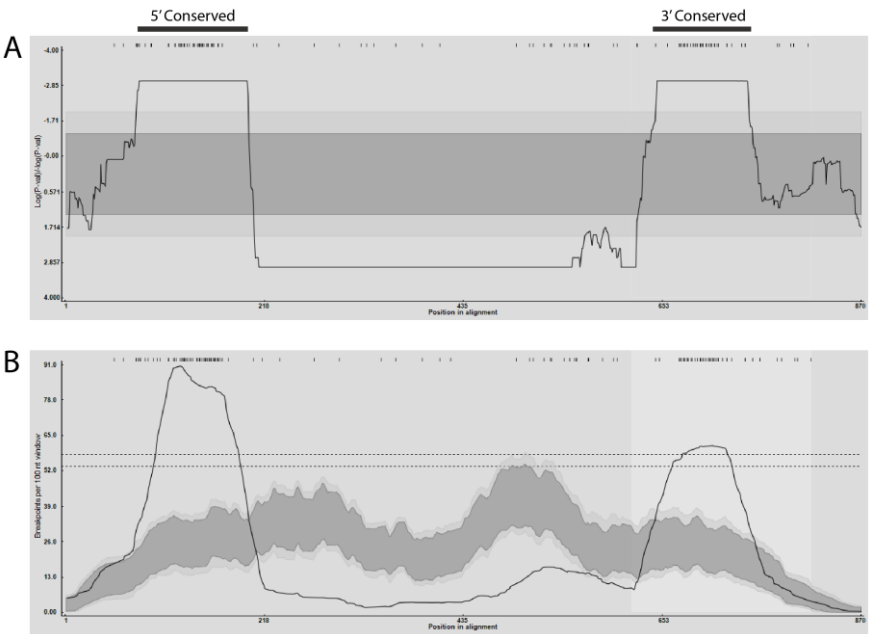


Figure 2. Predicted recombination events and their distribution among all *msp2/p44* genes of this study.(A) A breakpoint P-density plot of all predicted recombination events. (B) Predicted breakpoint site distribution. All 504 sequences are provided in Supplementary Table S1.

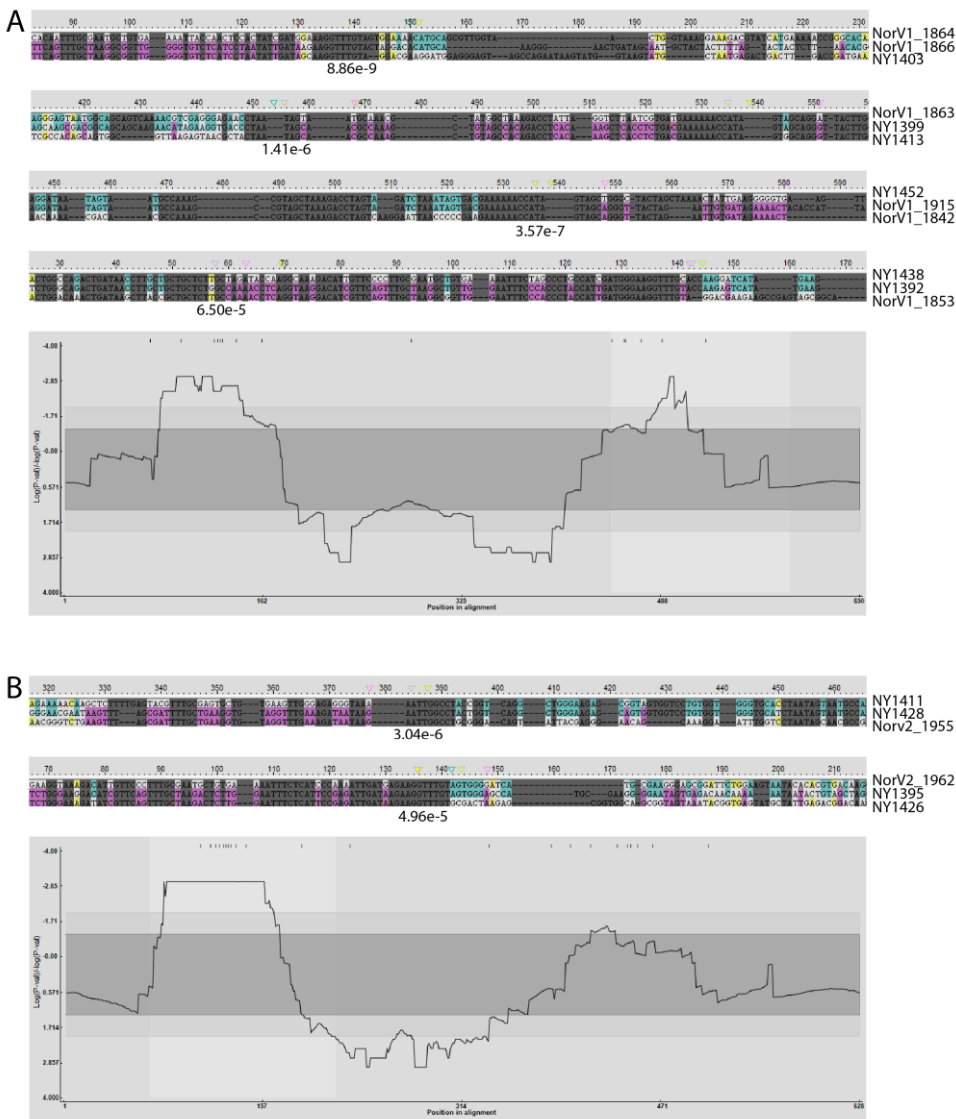


Figure 3. Predicted recombination events between individual HZ2_NY and NorShV1 or NorShV2 *msp2/p44* genes. Examples of recombination predicted with high probability between *msp2/p44* repertoires of (A) genes of the sheep strain-derived NorShV1 and human isolate-derived HZ2_NY, and (B) genes of the sheep strain-derived NorShV2 and HZ2_NY. Breakpoint P-density plots are provided beneath each of the examples for the larger alignments from which each example was extracted. The methods used and presentation of results are as described for Figure 1.

4. Discussion

This study demonstrates the outcomes of recombination events occurring between *msp2/p44* CVR gene regions that are largely isolate-specific. The circumstances under which these events occurred are unknown. Moreover, it is important to realize that it is not possible from these studies to ascribe sequences as being parental or recombinant in origin with certainty, as the evolutionary histories of these strains are unknown. From prior genomic analyses (18), however, it is clear that many USA strains infecting humans, dogs and horses from the Northeast and Mid-West are closely related. In these closely related strains recombination analysis suggests that initial recombination events are into the 5' and 3' conserved regions (hot-spots) flanking the CVR. In the more distantly related strains recombination events are more complex and can extend into the 3' variable region. The reasons for this polarity are not clear, but may be related to gene orientations relative to, and distances from, the origin of replication. In a prior repertoire analysis (20) that required >90% amino acid identities rather than >99% nucleotide identity as in the current study- a much lower threshold-

more repertoire genes were found to be shared. This suggests that point mutations as well as recombination in *msp2/p44* cause stepwise evolutionary change that can lead progressively to totally different surface antigen repertoires. It is not apparent from these analyses whether the mechanism of recombination among *msp2/p44* genes proceeds directly between unexpressed genes in the repertoire, via a multiple step mechanism involving genes present in the expression site specifically, or some measure of both. Gene conversion during antigenic variation normally involves a replacement of transcribed sequences in the expression site with duplicated sequences from the silent repertoire. However, at some much lower frequency it is likely this event, which is a form of DNA repair, may proceed in the reverse direction resulting in the insertion of novel sequence combinations into the silent repertoire.

There are several potential practical implications from the above analyses. First, unlike gene conversion of the *msp2/p44* expression site by different CVR's occurring during a single infection, such repertoire changes are expected to be more permanent and may facilitate the adaptation of the organism to different tick- and animal-host species. For example, the structures of the polypeptide sequences encoded by gene copies ApHZ2_NY1445 (minor parent) and ApWebster_WI2017 (recombinant; Figure 1A), are nearly identical as predicted by Robetta (<http://robetta.bakerlab.org/>; (21)) (Figure 4A). By contrast, the major parent, ApWebster_WI2064, differs in structure significantly (Figure 4B), with β -sheet and α -helix structure in what is a region of only random coil in the other two polypeptides. Recombination may result in novel structural combinations that could not only affect immune processing and recognition, but also the capacity of the protein to interact with alternative host components if incorporated into the expression site and expressed as MSP2 protein. As MSP2 is thought to have adhesin function (22) this may help to explain why *A. phagocytophilum* is currently an emerging disease of humans, yet has been known to infect sheep for >200 years (1). Second, there are epidemiologic implications for how an endemically stable situation, where most animals become persistently infected, may be rendered unstable by introduction of *A. phagocytophilum* strains with different repertoires. Observations consistent with this possibility have been made for infections of cattle caused by *A. marginale* (10, 23, 24). Third, the identification of less-favored intragenic sites for recombination may imply a requirement for conservation of structure in these regions of MSP2/P44 proteins, although given the diversity of sequences found there and the structural effects of these differences this would seem unlikely. Selection and retention for poor immunogenicity of the encoded polypeptide, and susceptibility to, and repair of, double-stranded DNA breaks during replication of these sequences are alternative possibilities that may also play roles. Fourth, orthologs of *msp2/p44* exist in other species taxonomically related to *Anaplasma* (the Pfam01617 family). For example, *Ehrlichia sp.* have 17-22 tandemly arranged members of this family (11) that are differentially expressed in situ rather than by recombination into a separate expression site (25-27). These genes express immunodominant surface antigens that have been considered targets for vaccination (28-30). It is possible that recombination among these gene copies may also occur and influence protective immune responses against heterologous strains. While it is not clear why this gene family is susceptible to DNA damage and rearrangements in some species but not others, this difference has ramifications for their suitability in immunization strategies. The work presented herein demonstrates the susceptibility of *A. phagocytophilum msp2/p44* to recombination within the silent *msp2/p44* repertoire outside of antigenic variation, and may help to explain both microbial persistence at the host population level and adaptation for infection of novel hosts.

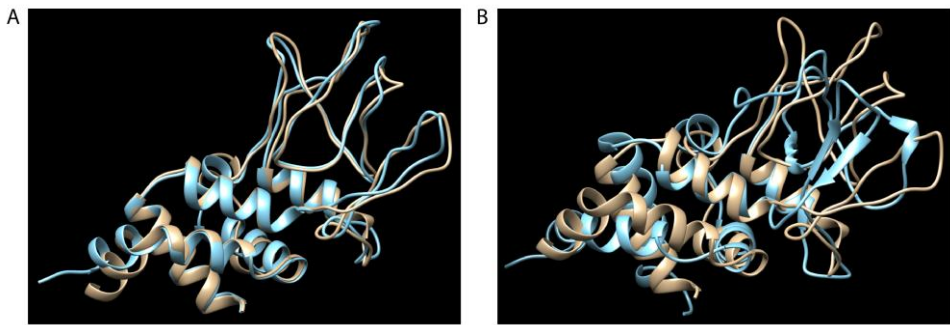


Figure 4. Superimposition of predicted structures of a recombinant MSP2 polypeptide and polypeptides encoded by its major and minor parent genes. (A) Structures of polypeptides encoded by genes HZ2_NY1445 (tan) and Webster_WI2017 (blue). (B) Structures of polypeptides encoded by genes HZ2_NY1445 (tan) and Webster_WI2064 (blue). In this example HZ2_NY1445 is a predicted recombinant gene, Webster_WI2017 is the minor parent, and Webster_WI2064 is the major parent.

Supplementary Materials: The following supporting information can be downloaded at the website of this paper posted on Preprints.org. Figure S1: Conserved and variable regions of HZ2_NY, Horse1_CA, and NorSh_V2 msp2/p44 repertoires; Figure S2: Alignment of all sequences used in this study; Table S1: Repertoire sequences of msp2/p44 genes within each of the *A. phagocytophilum* strains used in this analysis (.fasta format); Table S2: Statistical predictions for recombination among msp2/p44 repertoire sequences used in this study (.xlsx format).

Author Contributions: The authors contributed as follows: Conceptualization, A.B.; methodology, A.B. and D.A.; software, A.B. and D.A.; validation, A.B. and F.C.; formal analysis, A.B., D.A., and F.C.; investigation, A.B. and D.A.; writing—original draft preparation, A.B.; writing—review and editing, D.A. and F.C.; visualization, A.B. All authors have read and agreed to the published version of the manuscript except A.B., whose passage prior to completion precluded this.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: All sequences employed in this project are provided in Table S1, and are found in Genbank accessions #CP006616, #LANS00000000, #FLMF00000000, #FLMC00000000, #CP046639, and #CP015376.

Acknowledgments: The authors thank Joy, Mark, and David Barbet for their essential and unflagging support enabling the conduct of this project.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CVR	Central variable region of msp2/p44 genes
MSP2	Major surface protein 2

References

1. Stuen S. Anaplasma phagocytophilum - the most widespread tick-borne infection in animals in Europe. Vet Res Commun. 2007;31(Suppl. 1):79-84.

2. Brown WC, Barbet AF. Persistent Infections and immunity in ruminants to arthropod-borne bacteria in the family Anaplasmataceae. *Annu Rev Anim Biosci.* 2016;4:177-97.
3. Wang X, Kikuchi T, Rikihisa Y. Two monoclonal antibodies with defined epitopes of P44 major surface proteins neutralize *Anaplasma phagocytophilum* by distinct mechanisms. *Infect Immun.* 2006;74:1873-82.
4. Wang X, Rikihisa Y, Lai TH, Kumagai Y, Zhi N, Reed SM. Rapid sequential changeover of expressed p44 genes during the acute phase of *Anaplasma phagocytophilum* infection in horses. *Infect Immun.* 2004;72:6852-9.
5. Granquist EG, Stuenkel S, Crosby L, Lundgren AM, Alleman AR, Barbet AF. Variant-specific and diminishing immune responses towards the highly variable MSP2(P44) outer membrane protein of *Anaplasma phagocytophilum* during persistent infection in lambs. *Vet Immunol Immunopathol.* 2010;133:117-24.
6. Granquist EG, Stuenkel S, Lundgren AM, Bråten M, Barbet AF. Outer membrane protein sequence variation in lambs experimentally infected with *Anaplasma phagocytophilum*. *Infect Immun.* 2008;76:120-6.
7. Barbet AF, Meeus PF, Bélanger M, Bowie MV, Yi J, Lundgren AM, et al. Expression of multiple outer membrane protein sequence variants from a single genomic locus of *Anaplasma phagocytophilum*. *Infect Immun.* 2003;71:1706-18.
8. Lin Q, Rikihisa Y. Establishment of cloned *Anaplasma phagocytophilum* and analysis of p44 gene conversion within an infected horse and infected SCID mice. *Infect Immun.* 2005;73:5106-14.
9. Lin Q, Zhang C, Rikihisa Y. Analysis of involvement of the RecF pathway in p44 recombination in *Anaplasma phagocytophilum* and in *Escherichia coli* by using a plasmid carrying the p44 expression and p44 donor loci. *Infect Immun.* 2006;74:2052-62.
10. Lin Q, Ohashi N, Horowitz HW, Aguero-Rosenfeld ME, Raffalli J, Wormser GP, et al. Analysis of sequences and loci of p44 homologs expressed by *Anaplasma phagocytophilum* in acutely infected patients. *J Clin Microbiol.* 2002;40:2981-8.
11. Dunning Hotopp JC, Lin M, Madupu R, Crabtree J, Angiuoli SV, Eisen JA, et al. Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet.* 2006;2:e21.
12. Brayton KA, Kappmeyer LS, Herndon DR, Dark MJ, Tibbals DL, Palmer GH, et al. Complete genome sequencing of *Anaplasma marginale* reveals that the surface is skewed to two superfamilies of outer membrane proteins. *Proc Natl Acad Sci USA.* 2005;102:844-9.
13. Palmer GH, Bankhead T, Seifert HS. Antigenic variation in bacterial pathogens. *Microbiol Spectrum.* 2016;4:https://doi.org/10.1128/microbiolspec.vmbf-0005-2015.
14. Barbet AF, Lundgren AM, Yi J, Rurangirwa FR, Palmer GH. Antigenic variation of *Anaplasma marginale* by expression of MSP2 mosaics. *Infect Immun.* 2000;68:6133-8.
15. Futse JE, Brayton KA, Knowles DP, Palmer GH. Structural basis for segmental gene conversion in generation of *Anaplasma marginale* outer membrane protein variants. *Mol Microbiol.* 2005;57:212-21.
16. Rejmanek D, Foley P, Barbet AF, Foley J. Evolution of antigen variation in the tick-borne pathogen *Anaplasma phagocytophilum*. *Mol Biol Evol.* 2012;29:391-400.
17. Lin Q, Rikihisa Y, Ohashi N, Zhi N. Mechanisms of variable p44 expression by *Anaplasma phagocytophilum*. *Infect Immun.* 2003;71:5650-61.
18. Crosby FL, Eskeland S, Bø-Granquist EG, Munderloh UG, Price LD, Al-Khedery B, et al. Comparative whole genome analysis of an *Anaplasma phagocytophilum* strain isolated from Norwegian sheep. *Pathogens.* 2022;11:601.
19. Martin DP, Varsani A, Roumagnac P, Botha G, Maslamoney S, Schwab T, et al. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol.* 2021;7:veaa087.
20. Barbet AF, Al-Khedery B, Stuenkel S, Granquist EG, Felsheim RF, Munderloh UG. An emerging tick-borne disease of humans is caused by a subset of strains with conserved genome structure. *Pathogens.* 2013;2:544-55.
21. Kim DE, Chivian D, Baker D. Protein structure prediction and analysis using the Robetta server. *Nucleic Acids Research.* 2004;32 Suppl 2:W526-W31.
22. Park J, Choi KS, Dumler JS. Major surface protein 2 of *Anaplasma phagocytophilum* facilitates adherence to granulocytes. *Infection and Immunity.* 2003;71:4018-25.

23. Castañeda-Ortiz EJ, Ueti MW, Camacho-Nuez M, Mosqueda JJ, Mousel MR, Johnson WC, et al. Association of *Anaplasma marginale* strain superinfection with infection prevalence within tropical regions. *PLoS One*. 2015;10:e0120748.
24. Koku R, Futse JE, Morrison J, Brayton KA, Palmer GH, Noh SM. The use of the antigenically variable Major Surface Protein 2 in the establishment of superinfection during natural tick transmission of *Anaplasma marginale* in Southern Ghana. *Infect Immun*. 2023;91:e0050122.
25. Singu V, Liu H, Cheng C, Ganta RR. *Ehrlichia chaffeensis* expresses macrophage- and tick cell-specific 28-kilodalton outer membrane proteins. *Infect Immun*. 2005;73:79-87.
26. Singu V, Peddireddi L, Sirigireddy KR, Cheng C, Munderloh UG, Ganta RR. Unique macrophage and tick cell-specific protein expression from the p28/p30-outer membrane protein multigene locus in *Ehrlichia chaffeensis* and *Ehrlichia canis*. *Cell Microbiol*. 2006;8:1475-87.
27. Duan N, Ma X, Cui H, Wang Z, Chai Z, Yan J, et al. Insights into the mechanism regulating the differential expression of the P28-OMP outer membrane proteins in obligatory intracellular pathogen. *Emerg Microbes Infect*. 2021;10:461-71.
28. Nyika A, Barbet AF, Burridge MJ, Mahan SM. DNA vaccination with map1 gene followed by protein boost augments protection against challenge with *Cowdria ruminantium*, the agent of heartwater. *Vaccine*. 2002;20:1215-25.
29. Crocquet-Valdes PA, Thirumalapura NR, Ismail N, Yu X, Saito TB, Stevenson HL, et al. Immunization with *Ehrlichia* P28 outer membrane proteins confers protection in a mouse model of ehrlichiosis. *Clin Vaccine Immunol*. 2011;18:2018-25.
30. Budachetri K, Lin M, Chien RC, Zhang W, Brock GN, Rikihisa Y. Efficacy and immune correlates of OMP-1B and VirB2-4 vaccines for protection of dogs from tick transmission of *Ehrlichia chaffeensis*. *mBio*. 2022;13:e0214022.
31. Martin D, Rybicki E. RDP: detection of recombination amongst aligned sequences. *Bioinformatics*. 2000;16:562-3.
32. Salminen MO, Carr JK, Burke DS, McCutchan FE. Identification of breakpoints in intergenotypic recombinants of HIV type 1 by BOOTSCANing. *AIDS Research and Human Retroviruses*. 1995;11:1423-5.
33. Maynard Smith J. Analyzing the mosaic structure of genes. *Journal of Molecular Evolution*. 1992;34:126-9.
34. Posada D, Crandall KA. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. *Proc Natl Acad Sci USA*. 2001;98:13757-62.
35. Gibbs MJ, Armstrong JS, Gibbs AJ. Sister-Scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. *Bioinformatics*. 2000;16:573-82.
36. Weiller GF. Phylogenetic profiles: a graphical method for detecting genetic recombinations in homologous sequences. *Molecular Biology of Evolution*. 1998;15:326-35.
37. Holmes EC, Worobey M, Rambaut A. Phylogenetic evidence for recombination in Dengue virus. *Mol Biol Evol*. 1999;16:405.
38. Lam HM, Ratmann O, Boni MF. Improved algorithmic complexity for the 3SEQ recombination detection algorithm. *Mol Biol Evol*. 2018;35:247-51.
39. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, et al. UCSF Chimera- a visualization system for exploratory research and analysis. *Journal of Computational Chemistry*. 2004;25(13):1605-12.
40. Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. *Brief Bioinform*. 2019;20:1160-6.
41. Procter JB, Carstairs GM, Soares B, Mourão K, Ofoegbu TC, Barton D, et al. Alignment of biological sequences with Jalview. *Methods Mol Biol*. 2021;2231:203-24.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.