

Article

Comparative Genomics Identifies a Potential Marker of Human-Virulent *Anaplasma phagocytophilum*

Basima Al-Khedery¹ and Anthony F. Barbet^{1,2,*}

¹ Department of Infectious Diseases and Pathology, University of Florida, Gainesville, FL 32611, USA; E-Mail: balkhedery@ufl.edu

² Emerging Pathogens Institute, University of Florida, Gainesville, FL 32611, USA

* Author to whom correspondence should be addressed; E-Mail: barbet@ufl.edu; Tel.: +1-352-294-4119; Fax: +1-352-392-9704.

Received: 27 November 2013; in revised form: 18 December 2013 / Accepted: 7 January 2014 /

Published: 9 January 2014

Abstract: We have previously described a comparative genome analysis of nine strains of *Anaplasma phagocytophilum* that showed similarity between strains infecting humans and U.S. dogs and a more distant relationship with horse and ruminant strains. This suggested that it may be possible to distinguish human-infective strains using simple DNA sequence-based diagnostic tests. This would be of epidemiologic significance in identifying and tracking the presence of virulent strains in tick vector populations. Further analysis identified a gene that was present in several strains, including U.S. *Ap*-variant 1 (ruminant), MRK (horse), and European sheep, but was deleted in strains infecting U.S. humans and dogs, suggesting that it could be a useful marker of human virulence. A simple PCR test was developed to identify the presence/absence of this gene. The PCR test discriminated *A. phagocytophilum* strains from clinically affected humans and U.S. dogs from the strains more distantly related in genome sequence. This warrants further testing of globally diverse *A. phagocytophilum* strains to examine world-wide conservation of this gene.

Keywords: human anaplasmosis; *Anaplasma phagocytophilum*; comparative genomics; diagnosis; PCR

1. Introduction

Anaplasma phagocytophilum, a member of the order Rickettsiales, causes human fatalities in the U.S., Europe and Asia, and also infects multiple animal species. In the U.S., case reports increased from 348 in 2000 to 1,761 in 2010 [1], and the reported hospitalization rate is 36% [2]. Human anaplasmosis (HA) can be treated with antibiotics, but the symptoms, such as headache, fever, and muscle aches are non-specific and can be confused with other common diseases such as the flu, often leading to inappropriate therapy. Increasingly, there are reports of infections transmitted by blood transfusions in the U.S. and Europe [3,4]. The animal species that have been found infected include cattle, sheep, goats, horses, dogs, foxes, cats and rodents. Also, the tick vectors, such as *Ixodes scapularis* and *Ixodes pacificus* in the U.S., are known for the broad range of hosts on which they feed [5]. Different strains of *A. phagocytophilum* have different animal host predilections and not all strains infect all hosts [6]. This complex ecology has made it difficult to assess the risk of transmission to humans and institute control measures.

There is extensive genomic diversity within the *A. phagocytophilum* species. Numerous attempts have been made to link particular genotypes to host-tropism phenotypes with some, although limited, success. For example, in the U.S. a two-base difference in 16S ribosomal RNA has identified some strains infective to either ruminants (known as *Ap*-variant 1) or to mice (and it is thought to humans, which are known as *Ap*-ha strains) [7–9]. This method has insufficient discriminatory power, however. Further study of multiple strains worldwide has identified at least fifteen 16S variants, of which *Ap*-ha is one of the most common [10]. However, other variants have also been found in human infections and *Ap*-ha is not limited to humans. Moreover, there are many phenotypically untyped 16S variants in both the U.S. and Europe and multiple variants may co-exist in a single infection [11,12]. Similarly, methods based on polymorphisms in single genes such as *ankA* and *groEL* produce different strain clustering to one another and have failed to definitively categorize human-infective strains.

We recently completed a study analyzing high-throughput gene sequences of *A. phagocytophilum* strains from the U.S. and Europe [13]. The rodent, dog and human strains were similar to one another and to the previously sequenced human-infective strain HZ (98.79%–100% average genome nucleotide identity). The *Ap*-variant 1 strains were different in numerous regions (96.21%–96.28% average nucleotide identity to HZ). Here, we used comparative genomics to identify a gene deletion that has occurred in strains infecting humans and dogs in the U.S. but not in the more distantly related ruminant and horse strains. A simple PCR test was developed to identify the presence/absence of this gene. This marker may aid investigations of the spread by ticks of *A. phagocytophilum* strains which cause disease in humans.

2. Results and Discussion

Comparison of genomes by alignment (Figure 1A) showed lower nucleotide identities in *Ap*MRK (horse) and *Ap*NorV2 (sheep) compared to *Ap*HZ, *Ap*JM and *Ap*Dog, agreeing with our previous data [13]. The differences were found throughout the genomes, but were localized particularly to *mSP2/p44* pseudogenes close to the origin of replication (base #1 in a linear representation of the circular genome [13]). Additionally, we identified genome segments present only in some strains.

Figure 1B shows a region that was present in *ApMRK* and *ApNorV2* but was apparently deleted in the other three strains (demarcated with red arrows). This region is sandwiched between two pairs of inversely duplicated open reading frames (ORFs) annotated as APH_0919/APH_0920 and APH_0921/APH_0922 (encoding hypothetical proteins) in the CP000235 reference *ApHZ* genome (Figure 2 and Figure 3A). We initially identified two novel, relatively large, ORFs in this genome segment, one encoding a degenerate copy of the ABC transporter gene found elsewhere in all the genomes (annotated as APH_0986 in the CP000235 reference *ApHZ* genome) and a second encoding a gene with no known orthologs in other *A. phagocytophilum* strains, or in the most recent GenBank database. To verify that this was a gene deletion and not caused by misassembly, we compared Roche/454 reads from seven strains directly with this assembled region of *ApMRK* and *ApNorV2* (Figure 2). Clearly, there were reads encompassing this previously unidentified gene in both the *A. phagocytophilum* strains infecting Norwegian sheep, the U.S. ruminant *ApCRT35*, and the Californian horse (*ApMRK*) strain. There were no alignable reads from *ApHZ* (human), *ApJM* (rodent), or *ApDog*, indicating that this was indeed a gene deletion in those strains most closely related to the human-infective *ApHZ* strain.

Figure 1. Genome alignments comparing five strains of *A. phagocytophilum*. **(A)** The *ApMRK* and *ApNorV2* strains have numerous regions of lower identity compared to *ApHZ* than do the *ApDog* and *ApJM* strains. **(B)** There is also a significant genome deletion in *ApHZ*, *ApDog* and *ApJM* compared to *ApMRK* and *ApNorV2* strains. For each comparison row, the percent identities range from 50 to 100 from the bottom to the top of the rows. Panel B is a zoomed in region of panel A to show the deleted region. The top line in each panel indicates the reference genome, either *ApHZ* or *ApMRK*.

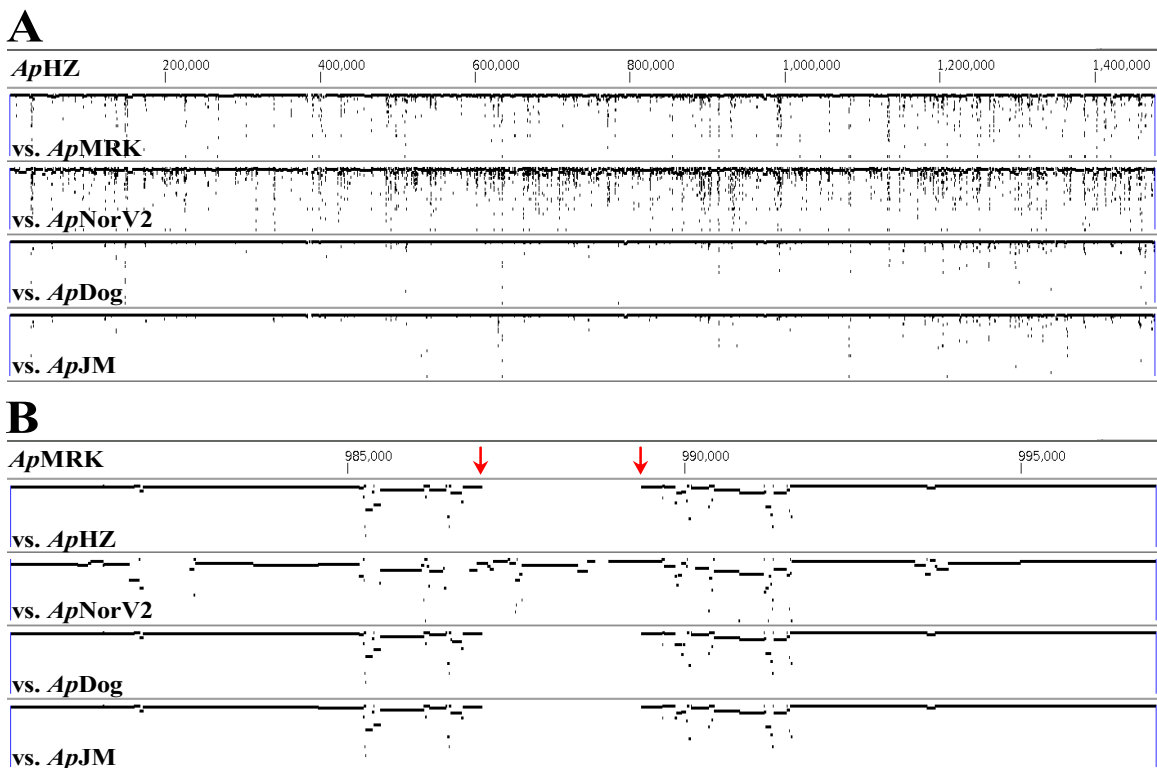
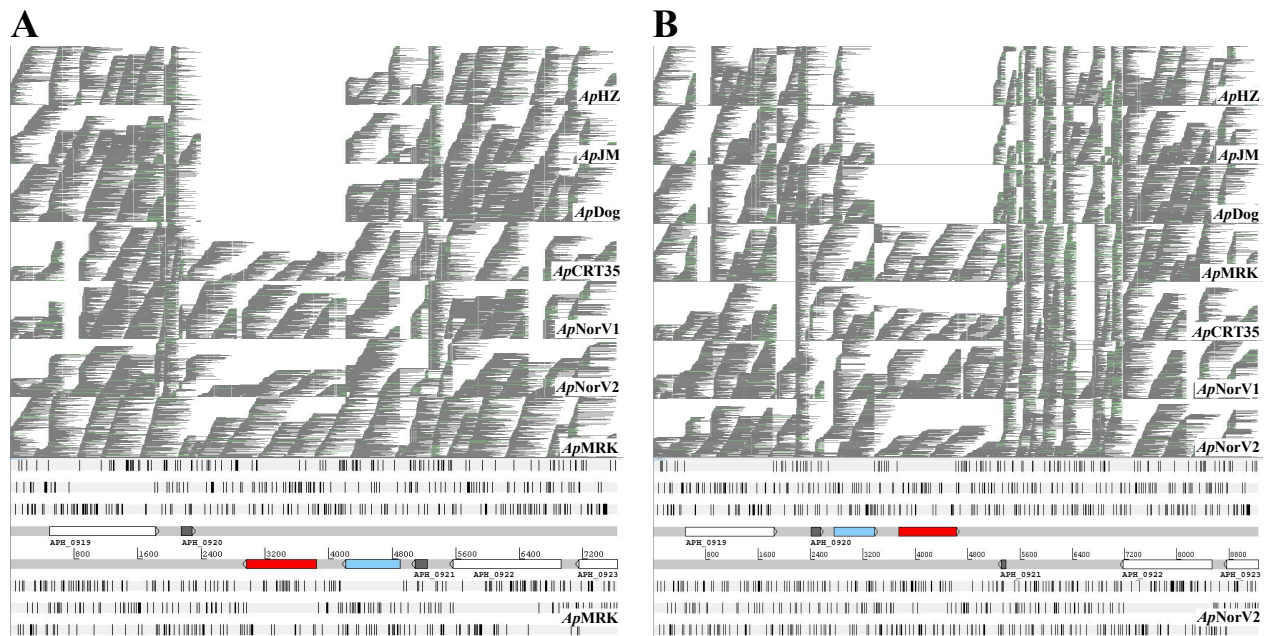


Figure 2. The deleted region encompasses the *drhm* gene (red). The extra copy of APH_0986 is shown in blue. Alignment of Roche/454 raw reads from seven *A. phagocytophilum* strains with the genome segment containing APH_0919 to APH_0923. (A) Alignment of reads with *ApMRK*; (B) Alignment of reads with *ApNorV2*.



We developed a PCR assay to investigate the presence or absence of this gene, termed *drhm* (for distantly related to human marker), in other *A. phagocytophilum* strains. The *drhm* gene was present in all U.S. strains tested that were previously identified as the ruminant-tropic *Ap*-variant 1, as well as present in U.S. Californian woodrat strains tested. It was absent in all cultured strains originally derived from human infections in Minnesota or Wisconsin and from blood taken from clinically infected humans in New York state (Figure 3B and 3C). The conserved *msp4* gene was used as an internal positive control to verify the presence of amplifiable *A. phagocytophilum* DNA in all samples.

This genome region appears to be prone to rearrangement, although not as frequent as in *msp2/p44* pseudogenes, as the *drhm* gene is maintained in *A. phagocytophilum* strains derived from Europe as well as from the U.S. Midwest and California. This is most evident from the finding that the deleted segment can occur in either orientation relative to its placement in *ApMRK*, as well as the existence of two opposing *drhm* genes flanking the degenerate ABC transporter gene copy in the *ApNorV1* genome (Figure 3A). It should be noted that no sequences related to the *drhm* gene could be detected in the vicinity of the full-length ABC transporter gene in these strains, nor at any other loci in the assembled *ApMRK* and *ApNorV2* contigs. The putative polypeptides encoded by *drhm* range in identity from 87 to 99% (94 to 99% identity at the nucleic acid level; data not shown) and were strongly predicted by different algorithms to be integral membrane proteins with 5–6 transmembrane (TM) segments (Figure 4).

Further analysis of the amino acid sequence by PSORT suggested a location as inner membrane proteins; some algorithms (e.g., SIGNALP V.2.0) predicted an N-terminal signal peptide from 1–37, but others (e.g., SIGNALP V.4.1) did not. Scanning for motifs at the PRINTS-S Protein Fingerprint

Database and other protein profile databases did not yield any significant hits suggestive of a potential function for the DRHM polypeptides.

Figure 3. The newly identified *drhm* gene could serve as a potential marker of human-virulent *A. phagocytophilum* strains. **(A)** Comparative maps from 7 *A. phagocytophilum* strains depicting a genomic region spanning two pairs of inversely duplicated ORFs, originally annotated as APH_0919/APH_0920 and APH_0921/APH_0922 in the human *ApHZ* strain reference genome (CP000235). Sequences corresponding to a degenerate duplication of the ABC transporter gene (APH_0986 in CP000235; blue arrows) and one to two copies of the herein identified *drhm* gene (red arrows) are also present in this region, but only in strains more distantly related in genome sequence to the human *ApHZ* strain. Maps are drawn to scale. For clarity, sequences related to the *ApHZ* APH_0920 and APH_0921 ORFs are aligned, and stippled lines are included to indicate gaps relative to the locus in the *ApNorV1* strain. **(B and C)** A PCR test developed to identify the presence/absence of the *drhm* gene clearly demonstrates the absence of this gene in human strain-related isolates. The human samples used in **(C)** are: 1, *ApWebster*; 2, *ApNY18*; 3, 4 and 5, New York patient clinical samples; 6, *ApMN1*; 7, *ApMN2*. All template DNAs were obtained from *in vitro* cultures except for *ApNorV1*, *ApNorV2*, New York patient clinical samples, Swedish dog and California woodrat samples.

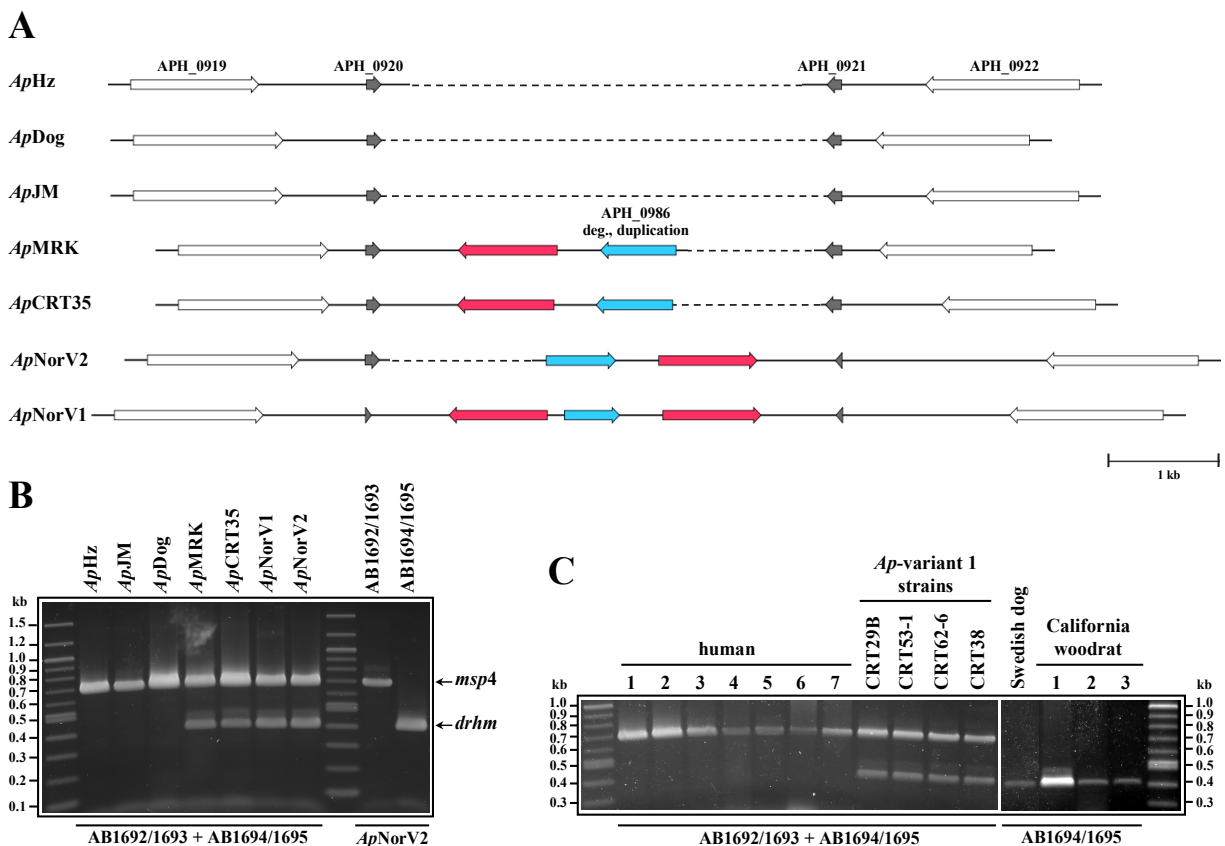
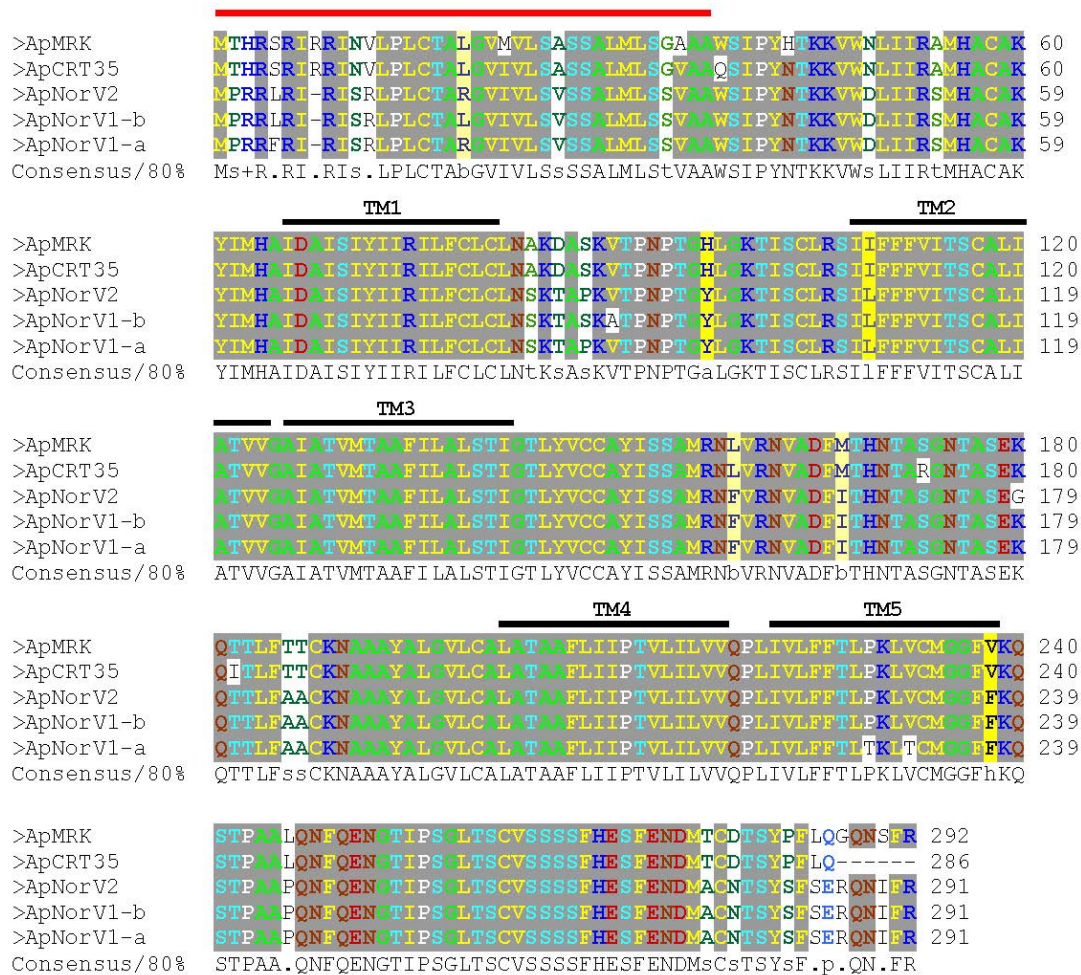


Figure 4. Multiple sequence alignment of amino acid sequences encoded by *drhm* from different strains of *A. phagocytophilum*. The positions of a putative signal peptide (red line) and five predicted TM segments (black lines) are shown.



The high worldwide prevalence of *A. phagocytophilum*, the continuing spread of its vector *Ixodes* ticks into new geographic areas [14], the potential for severe human disease and contamination of blood supplies make this emerging infection a cause for concern. We have shown data suggesting that, at least in the U.S., the presence of the *drhm* gene is indicative of strains that are not infecting humans but are more closely related to the long-known ruminant-infective strains. However, currently the number of strains analyzed at the whole genome level is small and should be expanded to verify this interpretation. One must also consider alternative possibilities. For example, that the gene deletion represents an artifact from adaptation of strains to *in vitro* culture conditions. We believe this to be less likely as the deletion was present in *A. phagocytophilum* DNA derived directly from clinical samples. It is possible that milder human infections may be caused by other strains but have not yet been recognized. Interestingly, in the U.S. the strains identified that cause disease in humans and dogs are similar at the whole genome level. Some data suggest this may not be true globally; *A. phagocytophilum* strains infecting Swedish dogs have not grouped phylogenetically with U.S. dog strains [15]. Similarly, our preliminary data shows that a strain from an infected Swedish dog

possesses the *drhm* gene (Figure 3C), unlike the U.S. dog strains analyzed (Figure 3B, and data not shown). There is a significantly lower number of human infections described in Europe compared to the U.S. for reasons that are not totally clear and are likely due to various factors such as levels of tick exposure, infection prevalence and disease surveillance. It is also possible that those strains that are most virulent for humans are not widely distributed in the tick vector and result from recent evolutionary divergence. This could produce localized foci of human infections, similar to those observed. Comparative genomics and the *drhm* PCR test will provide an initial rapid classification of *A. phagocytophilum* strains derived from animal reservoirs and ticks that can be investigated further for human virulence.

3. Experimental Section

3.1. *A. phagocytophilum* Strain Origins and Genomic DNA Isolation

The origins of the *A. phagocytophilum* strains used in this study are as follows: *ApHZ*, human, Westchester, New York, USA [16]; *ApNY18*, human, New York, USA [17]; *ApWebster*, human, Wisconsin [18]; *ApMN1* and *ApMN2*, human, Minnesota; *ApDog*, dog, Minnesota; *ApJM*, meadow jumping mouse (*Zapus hudsonius*), Camp Ripley, Minnesota; *ApMRK*, horse, California [19,20]; *Ap-CRT35*, -*CRT29B*, -*CRT38*, -*CRT53-1* and -*CRT62-6* (*Ap*-variant 1 strains, Camp Ripley tick (*I. scapularis*), Minnesota [21]; *ApNorV1* and *ApNorV2*, sheep, Norway [13,22]. *ApDog*, *ApCRT35*, *ApNorV1* and *ApNorV2* were previously defined as *ApDog2*, *ApVar-1*, *ApNorLamb-V1* and *ApNorLamb-V2*, respectively [22].

Strains from *in vitro* culture included: *ApHZ*, *ApNY18*, *ApMN1*, *ApMN2*, *ApWebster*, *ApJM* and *ApMRK*, propagated in HL-60 cells, and *ApDog* and *ApCRT35*, maintained in the *I. scapularis* ISE6 tick cell line. Starter cultures of *ApHZ*, *ApJM*, and *ApMRK* were generously provided by Dr. Ulrike G. Munderloh, as was genomic DNA (gDNA) from *ApDog*, *ApCRT35*, *ApMN1* and *ApMN2*. The three New York patient samples were generously provided by Dr. Susan Wong, NY State Dept. of Health [23], *ApNorV1* and *ApNorV2* by Drs. Snorre Stuen and Erik G. Granquist, and the *Ap*-variant 1 strains by Dr. Robert F. Massung. *A. phagocytophilum* gDNA was prepared as described previously [22]. Genomic DNA from California dusky-footed woodrats (*Neotoma fuscipes*) was kindly provided by Dr. Janet E. Foley, and from the Swedish dog by Dr. Anneli Bjöersdorff.

3.2. Ethics Statement

The experimental study in sheep was approved by the Norwegian Animal Research Authority.

3.3. 454 Genome Sequencing and Bioinformatics

Genomic DNA was sequenced on the Roche/454 Genome Sequencer as previously described [13,22], with genome coverage ranging from 31.3X to 72.1X. Briefly, regular read libraries were generated for *ApHZ*, *ApDog*, *ApJM*, *ApMRK*, *Ap-CRT35*, *ApNorV1* and *ApNorV2*. Additionally, 3 kb paired end libraries were made for *ApHZ* and *ApMRK*. Genome drafts were assembled using the CLC Genomics Workbench software suite (version 4.0–4.9) with default parameters: length fraction, 0.5; similarity, 0.8; and for paired end reads, minimum distance, 1,500/maximum distance, 4,500. Gapped drafts for the

ApMRK and *ApNorV2* genomes were generated by a combination of mapping and *de novo* assembly. Briefly, initial consensus sequences for each strain were obtained by mapping the respective reads against the fully annotated Sanger sequenced *ApHZ* genome (GenBank CP000235). Regions with corresponding *de novo* contigs were manually identified and replaced with *de novo* contig sequences. Reads were again mapped to the resulting consensus sequences and underlying aligned reads were inspected for conflicts and gaps, which were manually corrected as described [13,22]. In this fashion, 9 finalized contigs were obtained for *ApMRK*, and 23 contigs for *ApNorV2*. Residual gaps correspond mainly to some *msp2/p44* gene clusters, the large R3/R4 repeat regions of the *virB6-4* gene [22], and large genome duplications/insertions. To perform the analysis presented in Figure 1, for each of these two strains, a contiguous consensus genome sequence was generated consisting of the finalized contigs joined by Ns and ordered according to the CP000235 reference genome. To obtain the loci depicted in Figure 3A for *ApCRT35* and *ApNorV1*, consensus sequences were generated by mapping the respective reads against the corresponding regions in the *ApMRK* and *ApNorV2* genome drafts, and conflicts and gaps were manually resolved as above.

Genome alignments (Figure 1) were conducted using MUGSY and displayed with GMAJ [24]. To align Roche/454 reads with assembled contigs (Figure 2) LASTZ running on a local instance of GALAXY [25–27] on the University of Florida high performance computer cluster was used (75% identity cutoff). The SAM-format output files were converted to BAM, sorted and indexed with SAMTOOLS [28] and displayed with ARTEMIS [29,30]. Amino acid sequences were aligned with MAFFT [31] and displayed with CHROMA [32]. The putative signal peptide region and TM segments highlighted in Figure 4 were predicted with SIGNALP and PSORT [33–36].

3.4. PCR Amplification

PCR products spanning *msp4* and the *drhm* gene (~680 bp and 390 bp, respectively; Fig 3B and 3C) were derived in a single reaction or separately using Takara's PrimeSTAR GXL DNA Polymerase system (Clontech Laboratories, Mountain View, CA, USA). Reactions contained 5 ng gDNA or 5–12 µL from very low concentration samples, and: 1.25 U polymerase, 1.0 mM MgCl₂, 200 µM each dNTP, and 200 nM each primer, in total 50 µL. Universal primers for *msp4* were: AB1692, 5'-TAATGATGCGTCTGATGTTAGCG-3', forward; AB1693, 5'-CACCCACCTGCTATGTTTACACG-3', reverse, and for *drhm*: AB1694, 5'-TATCTTAGCTCTCTCCACCATAG-3', forward; AB1695, 5'-AACTAGACGATGATACACAAGATG-3', reverse. Following the manufacturer's recommendations, 3-step PCR was performed with 30 cycles of 10 s denaturing at 98 °C, 15 s annealing at 60 °C and 90 s extension at 68 °C. Final extension was for 5 min at 72 °C. PCR products were analyzed on 1.5% agarose gels alongside a 100 bp DNA ladder (New England Biolabs, Beverly, MA, USA).

3.5. GenBank Accession Numbers

GenBank accession numbers for the *drhm* gene in *ApMRK*, *ApCRT35*, *ApNorV1* (*drhm-a*), *ApNorV1* (*drhm-b*) and *ApNorV2* are KF905325, KF905326, KF905327, KF905328 and KF905329, respectively.

4. Conclusions

These data showed significant differences between *A. phagocytophilum* genomes derived from different host animals and geographic locations. Comparative genomics revealed a consistent gene deletion in the strains infecting humans in the U.S. compared to many other global *A. phagocytophilum* strains. A PCR assay was developed to detect the presence of this gene, *drhm*. This could be applied rapidly for strain identification in epidemiological studies to track *A. phagocytophilum* strains with varying degrees of potential infectivity to humans.

Acknowledgments

We thank Anna M. Lundgren for expert technical assistance with PCR analysis, and Savita Shanker for high-throughput DNA sequencing. This investigation received support from grants RO1 GM081714 and GM081714-03S1.

Conflicts of Interest

The authors declare no conflict of interest.

References

1. CDC. Statistics and Epidemiology. Annual Cases of Anaplasmosis in the United States. Available online: <http://www.cdc.gov/anaplasmosis/stats/> (accessed on 5 September 2013).
2. Dahlgren, F.S.; Mandel, E.J.; Krebs, J.W.; Massung, R.F.; McQuiston, J.H. Increasing incidence of *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* in the United States, 2000–2007. *Am. J. Trop. Med. Hyg.* **2011**, *85*, 124–131.
3. Annen, K.; Friedman, K.; Eshoa, C.; Horowitz, M.; Gottschall, J.; Straus, T. Two cases of transfusion-transmitted *Anaplasma phagocytophilum*. *Am. J. Clin. Pathol.* **2012**, *137*, 562–565.
4. Jereb, M.; Pecaver, B.; Tomazic, J.; Muzlovic, I.; Avsic-Zupanc, T.; Premru-Srsen, T.; Levicnik-Stezinar, S.; Karner, P.; Strle, F. Severe human granulocytic anaplasmosis transmitted by blood transfusion. *Emerg. Infect. Dis.* **2012**, *18*, 1354–1357.
5. Jin, H.; Wei, F.; Liu, Q.; Qian, J. Epidemiology and control of human granulocytic anaplasmosis: A systematic review. *Vector Borne Zoonotic Dis.* **2012**, *12*, 269–274.
6. Foley, J.E.; Nieto, N.C.; Massung, R.; Barbet, A.; Madigan, J.; Brown, R.N. Distinct ecologically relevant strains of *Anaplasma phagocytophilum*. *Emerg. Infect. Dis.* **2009**, *15*, 842–843.
7. Massung, R.F.; Priestley, R.A.; Miller, N.J.; Mather, T.N.; Levin, M.L. Inability of a variant strain of *Anaplasma phagocytophilum* to infect mice. *J. Infect. Dis.* **2003**, *188*, 1757–1763.
8. Massung, R.F.; Mather, T.N.; Priestley, R.A.; Levin, M.L. Transmission efficiency of the Ap-variant 1 strain of *Anaplasma phagocytophilum*. *Ann. N. Y. Acad. Sci.* **2003**, *990*, 75–79.
9. Massung, R.F.; Mather, T.N.; Levin, M.L. Reservoir competency of goats for the Ap-variant 1 strain of *Anaplasma phagocytophilum*. *Infect. Immun.* **2006**, *74*, 1373–1375.
10. Rar, V.; Golovljova, I. *Anaplasma*, *Ehrlichia*, and “Candidatus neoehrlichia” Bacteria: Pathogenicity, biodiversity, and molecular genetic characteristics, a review. *Infect. Genet. Evol.* **2011**, *11*, 1842–1861.

11. Stuen, S.; van de Pol, I.; Bergström, K.; Schouls, L.M. Identification of *Anaplasma phagocytophila* (formerly *Ehrlichia phagocytophila*) variants in blood from sheep in Norway. *J. Clin. Microbiol.* **2002**, *40*, 3192–3197.
12. Stuen, S.; Nevland, S.; Moum, T. Fatal cases of tick-borne fever (tbf) in sheep caused by several 16S rRNA gene variants of *Anaplasma phagocytophilum*. *Ann. N. Y. Acad. Sci.* **2003**, *990*, 433–434.
13. Barbet, A.; Al-Khedery, B.; Stuen, S.; Granquist, E.; Felshein, R.; Munderloh, U. An emerging tick-borne disease of humans is caused by a subset of strains with conserved genome structure. *Pathogens* **2013**, *2*, 544–555.
14. Patz, J.A.; Olson, S.H.; Uejio, C.K.; Gibbs, H.K. Disease emergence from global climate and land use change. *Med. Clin. North Am.* **2008**, *92*, 1473–1491.
15. Morissette, E.; Massung, R.F.; Foley, J.E.; Alleman, A.R.; Foley, P.; Barbet, A.F. Diversity of *Anaplasma phagocytophilum* strains, USA. *Emerg. Infect. Dis.* **2009**, *15*, 928–931.
16. Rikihisa, Y.; Zhi, N.; Wormser, G.P.; Wen, B.; Horowitz, H.W.; Hechemy, K.E. Ultrastructural and antigenic characterization of a granulocytic ehrlichiosis agent directly isolated and stably cultivated from a patient in New York state. *J. Infect. Dis.* **1997**, *175*, 210–213.
17. Aguero-Rosenfeld, M.E.; Horowitz, H.W.; Wormser, G.P.; McKenna, D.F.; Nowakowski, J.; Muñoz, J.; Dumler, J.S. Human granulocytic ehrlichiosis: A case series from a medical center in New York state. *Ann. Intern. Med.* **1996**, *125*, 904–908.
18. Asanovich, K.M.; Bakken, J.S.; Madigan, J.E.; Aguero-Rosenfeld, M.; Wormser, G.P.; Dumler, J.S. Antigenic diversity of granulocytic ehrlichia isolates from humans in Wisconsin and New York and a horse in California. *J. Infect. Dis.* **1997**, *176*, 1029–1034.
19. Gribble, D.H. Equine ehrlichiosis. *J. Am. Vet. Med. Assoc.* **1969**, *155*, 462–469.
20. Madigan, J.E.; Gribble, D. Equine ehrlichiosis in northern California: 49 cases (1968–1981). *J. Am. Vet. Med. Assoc.* **1987**, *190*, 445–448.
21. Massung, R.F.; Levin, M.L.; Munderloh, U.G.; Silverman, D.J.; Lynch, M.J.; Gaywee, J.K.; Kurtti, T.J. Isolation and propagation of the Ap-variant 1 strain of *Anaplasma phagocytophilum* in a tick cell line. *J. Clin. Microbiol.* **2007**, *45*, 2138–2143.
22. Al-Khedery, B.; Lundgren, A.M.; Stuen, S.; Granquist, E.G.; Munderloh, U.G.; Nelson, C.M.; Alleman, A.R.; Mahan, S.M.; Barbet, A.F. Structure of the Type IV secretion system in different strains of *Anaplasma phagocytophilum*. *BMC Genomics* **2012**, *13*, 678.
23. Barbet, A.F.; Meeus, P.F.; Bélanger, M.; Bowie, M.V.; Yi, J.; Lundgren, A.M.; Alleman, A.R.; Wong, S.J.; Chu, F.K.; Munderloh, U.G.; *et al.* Expression of multiple outer membrane protein sequence variants from a single genomic locus of *Anaplasma phagocytophilum*. *Infect. Immun.* **2003**, *71*, 1706–1718.
24. Angiuoli, S.V.; Salzberg, S.L. Mugsy: Fast multiple alignment of closely related whole genomes. *Bioinformatics* **2011**, *27*, 334–342.
25. Goecks, J.; Nekrutenko, A.; Taylor, J. Galaxy: A comprehensive approach for supporting accessible, reproducible, and transparent computational research in the life sciences. *Genome Biol.* **2010**, *11*, R86.
26. Blankenberg, D.; von Kuster, G.; Coraor, N.; Ananda, G.; Lazarus, R.; Mangan, M.; Nekrutenko, A.; Taylor, J. Galaxy: A web-based genome analysis tool for experimentalists. *Curr. Protoc. Mol. Biol.* **2010**, Chapter 19, Unit 19.10.1–Unit 19.10.21.

27. Giardine, B.; Riemer, C.; Hardison, R.C.; Burhans, R.; Elnitski, L.; Shah, P.; Zhang, Y.; Blankenberg, D.; Albert, I.; Taylor, J.; *et al.* Galaxy: A platform for interactive large-scale genome analysis. *Genome Res.* **2005**, *15*, 1451–1455.
28. Li, H.; Handsaker, B.; Wysoker, A.; Fennell, T.; Ruan, J.; Homer, N.; Marth, G.; Abecasis, G.; Durbin, R. The sequence alignment/map format and samtools. *Bioinformatics* **2009**, *25*, 2078–2079.
29. Carver, T.; Harris, S.R.; Berriman, M.; Parkhill, J.; McQuillan, J.A. Artemis: An integrated platform for visualization and analysis of high-throughput sequence-based experimental data. *Bioinformatics* **2012**, *28*, 464–469.
30. Carver, T.; Harris, S.R.; Otto, T.D.; Berriman, M.; Parkhill, J.; McQuillan, J.A. Bamview: Visualizing and interpretation of next-generation sequencing read alignments. *Brief. Bioinform.* **2013**, *14*, 203–212.
31. Katoh, K.; Toh, H. Recent developments in the mafft multiple sequence alignment program. *Brief. Bioinform.* **2008**, *9*, 286–298.
32. Goodstadt, L.; Ponting, C.P. Chroma: Consensus-based colouring of multiple alignments for publication. *Bioinformatics* **2001**, *17*, 845–846.
33. Emanuelsson, O.; Brunak, S.; von Heijne, G.; Nielsen, H. Locating proteins in the cell using targetp, signalp and related tools. *Nat. Protoc.* **2007**, *2*, 953–971.
34. Nielsen, H.; Engelbrecht, J.; Brunak, S.; von Heijne, G. Identification of prokaryotic and eukaryotic signal peptides and prediction of their cleavage sites. *Protein Eng.* **1997**, *10*, 1–6.
35. Petersen, T.N.; Brunak, S.; von Heijne, G.; Nielsen, H. Signalp 4.0: Discriminating signal peptides from transmembrane regions. *Nat. Methods* **2011**, *8*, 785–786.
36. Nakai, K.; Horton, P. Psort: A program for detecting sorting signals in proteins and predicting their subcellular localization. *Trends Biochem. Sci.* **1999**, *24*, 34–36.

© 2014 by the authors; licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution license (<http://creativecommons.org/licenses/by/3.0/>).