JOURNAL OF CLINICAL MICROBIOLOGY, July 2007, p. 2312–2315 Vol. 45, No. 7
0095-1137/07/$08.00+0 doi:10.1128/JCM.00704-07

Contribution of Polymorphisms in *ankA*, *gltA*, and *groESL* in Defining Genetic Variants of *Anaplasma phagocytophilum*▽†

Sanjay K. Shukla,[1]* Vijay Aswani,[2] Patrick J. Stockwell,[1] and Kurt D. Reed[1]

Molecular Microbiology Laboratory, Marshfield Clinic Research Foundation,[1] and Department of Internal Medicine and Pediatrics, Marshfield Clinic,[2] Marshfield, Wisconsin 54449

Received 30 March 2007/Returned for modification 2 May 2007/Accepted 10 May 2007

Analysis of several nucleotide polymorphisms in polymorphic genes (*ankA*, *gltA*, and *groESL*) from 16S rRNA gene-based genetic variants of *Anaplasma phagocytophilum* from dogs in the western United States defined at least two sets of multigene polymorphisms to further characterize these variants. The multigene polymorphism approach holds promise for development of a genotyping scheme for this important pathogen.

Anaplasma phagocytophilum is a tick-borne pathogen that causes granulocytic anaplasmosis in a number of mammals, cats, llamas, including dogs, horses, and humans (3, 5, 16, 23, 25, 33). Strains of *A. phagocytophilum* show disparities in clinical severity, disease manifestation, reservoir competency, and antigenic diversity (2, 6, 19–22, 27), suggesting the importance of studying the pathogen's clonal structure and population genetics. One of the most frequently used genes for phylogenetic studies has been the 16S rRNA gene because of a slow rate of mutation in its variable intragenic region interspersed with stable constant domains (24) and the availability of a large sequence database in GenBank for sequence comparison. A small number of genetic variants based on polymorphisms in the 16S rRNA gene of *A. phagocytophilum* from humans, horses, sheep, llamas, white-tailed deer, and ticks have been described (3, 4, 10, 19, 22, 27). Their biological significance has also been discussed (20–22, 27). We recently described five genetic variants of *A. phagocytophilum* from seven dogs in Washington State that had clinical diagnoses of granulocytic anaplasmosis. This included five cases of coinfection with more than one variant (22). In addition to the 16S rRNA gene, the roles of *ankA* (a gene coding for a cytoplasmic protein antigen), *gltA* (the gene for a citrate synthase), and *groESL* (a gene coding for a heat shock protein) of *A. phagocytophilum* strains from different hosts have been explored to determine their phylogeny (7, 18, 26, 29, 32). In this investigation, we explored the combined contribution of polymorphisms in these three phylogenetically informative genes in order to characterize genetic variants and better define the clonal structure of this pathogen.

(This study was presented in part at the 5th Ticks and Tick-Borne Pathogens Meeting in Neuchatel, Switzerland, August/September 2005.)

Each dog in this study was a resident of Washington State. Their travel histories, clinical syndromes, pathology laboratory findings, and evidence of coinfection by one or more variants of *A. phagocytophilum* based on the polymorphisms in the 16S rRNA gene have been described previously (22). DNA was extracted from the EDTA blood samples from seven dogs that were infected with known anaplasmal variants. Portions of *ankA*, *gltA*, and *groESL* were amplified using the following primer sets: *ankA* primers LA6/LA1 (8), *gltA* primers HGM28F/HG1257R (13), and *groESL* primers HS1/HS6 (9, 28). Two or three independent amplicons were generated for each gene, and on average seven independent sequencing reactions were performed, covering both sense and antisense strands. Therefore, each identified polymorphism was the output from, on average, a sevenfold sequence coverage of the relevant portions of all three genes after being compared to all the available sequences in GenBank. To maintain consistency with the published literature, polymorphisms in *ankA*, *gltA*, and *groESL* have been presented in the context of GenBank accession numbers AF100885, AF304136, and U96728, respectively. Novel sets of polymorphisms in each of the three genes were identified, and individual sets were designated as alpha, beta, or gamma sets to provide clarity. The use of the term "variant" was restricted to novel strains based on unique 16S rRNA gene polymorphism criteria only.

Phylogenetic relatedness of each variant with that of the available sequences from GenBank was determined by phylogenetic analysis. All sequences were aligned, and a neighbor-joining tree was built using the minimum-evolution criterion. Based on the structure of the tree, representative taxa from the significant clades were selected for further analysis. Since these were protein-coding genes, each alignment was translated using MacClade (17) and checked for premature termination codons and inappropriate gaps or alignments. LogDet and maximum-likelihood distance trees were generated using PAUP* 4.0b10 (30).

Nucleotide polymorphism sets. Identified nucleotide polymorphisms in a single copy of each of three genes are summarized in Tables 1 to 3. The unique polymorphism profile for each gene was designated to represent a polymorphism set. In the context of this study, a polymorphism set is expected to be specific to strains representing each genetic variant. For *ankA*, two polymorphism sets (alpha and beta) were identified in six of the seven blood samples. The *ankA* alpha set has the

* Corresponding author. Mailing address: Molecular Microbiology Laboratory, Marshfield Clinic Research Foundation, 1000 North Oak Avenue, Marshfield, WI 54449. Phone: (715) 389-5363. Fax: (715) 389-3808. E-mail: shukla.sanjay@mcrf.mfldclin.edu.

† Supplemental material for this article may be found at http://jcm.asm.org/.

▽ Published ahead of print on 16 May 2007.

TABLE 1. Unique sets of polymorphisms in the *ankA* genes associated with *A. phagocytophilum* variants

GenBank no.[a] or set	Nucleotide(s) at position:										
	2288	2290	2292	2323	2329	2330	2335	2377	2379	2443	2620
AF100885	G	A	G	C	T	G	G	A	T	TTT	A
Alpha	A	T	A	G	C	T	C	A	G	None[b]	G
Beta	A	T	A	G	C	T	C	C	G	None[b]	G

[a] GenBank accession number used to identify nucleotide positions.
[b] Deletion.

TABLE 3. Unique sets of polymorphisms in the *groESL* genes associated with *A. phagocytophilum* variants

GenBank no.[a] or set	Nucleotide at position:			
	154	285	585	939
U96728	G	G	T	T
Alpha	G	A	T	C
Beta	G	A	C	C
Gamma	A	A	T	C

[a] GenBank accession number used to identify nucleotide positions.

following polymorphisms at the indicated nucleotide positions (underlined): AGG2288A (Arg→Lys), ATG2290/2292TTA (Met→Leu), CCA2323G (Pro→Ala), TGT2329-2330CTT (Cys→Leu), GAT2335C (Asp→His), AAT2379G (Asn→Lys), TTT deletion at 2443 (loss of Phe) and ACT2620G (Thr→Ala). The *ankA* beta set has all the polymorphisms of the alpha set plus the AAT2377/2379CAG (Asn→Gln) change. The 3-bp deletion (TTT) seen in the *ankA* gene has also been seen in strains from the upper Midwest (18). For *gltA*, two polymorphism sets were identified. The *gltA* alpha set had the AAG267T (Lys→Asn) and GTG314→C (Val→Ala) polymorphisms, whereas the beta set had the following polymorphisms in addition to the one described for the *gltA* alpha set: GTA269A (Val→Ile), ACG66A (synonymous change), GAC598*A* (Asp→Asn), and GTG696A (synonymous change). Three unique *groESL* sets of polymorphisms were identified. The *groESL* alpha set has mutations at the following nucleotide positions: ACG285A (synonymous change) and GCT939C (synonymous change). The *groESL* beta set has the GAT585C synonymous change in addition to the alpha set of polymorphisms. The gamma set has the ATG154A (Met→Ile) change as well as the polymorphisms described in the alpha set. Dogs 4 and 5 were coinfected with strains harboring both alpha and beta sets, whereas dog 6 was infected with strains harboring alpha and gamma sets. The remaining dogs were infected with the strain harboring the alpha sets only. The multiset of polymorphisms in each of the three genes is consistent with our previously reported observation that some of these dogs were coinfected with more than one variant (22). The premise was that the *A. phagocytophilum* genome harbors only a single copy of these genes and therefore any polymorphisms/heterozygosity in these genes would reflect coinfections with more than one genotype or variant.

A tentative genotype scheme from the summary of the combination of all the polymorphism sets in *ankA*, *gltA*, and *groESL* including our previously published data for 16S rRNA gene (22) for individual dog samples is presented in Table 4. The above genotyping criteria were based on the following parameters: (i) molecular characteristics of the Washington (WA) variants described earlier (22) and (ii) the presence and absence of individual sets of polymorphisms (alpha, beta, and gamma sets) in the three genes in different dog samples. We also had to make the assumptions that we had identified all the possible variants in each sample by the 16S rRNA gene sequencing approach (22) and identified all the polymorphism sets for each of the three genes. In addition, if any particular set of polymorphisms in a gene was present in all samples, we assumed they represented the same variant. For example, dog 5, which was infected with WA variant 3, also harbored beta sets of polymorphisms of *ankA* and *gltA* plus either the *groESL* alpha or beta set. Similarly, WA variant 4, which was present in dogs 2, 3, 4, 6, 7, and 8, was further defined by the presence of alpha sets of polymorphisms in the *ankA*, *gltA*, and *groESL* genes. Due to the limitation of this approach, we were not able to identify polymorphisms in *ankA*, *gltA*, and *groESL* genes of WA variants 1, 2, and 5. The data shown in Table 4 enabled us to propose a genotype of at least two variants of WA variants 3 and 4. According to this scheme, WA variant 3 harbors the *ankA* beta set, the *gltA* beta set, and either the alpha or beta *groESL* set, whereas WA variant 4 harbors the *ankA* alpha set, *gltA* alpha set, and *groESL* alpha set.

Phylogenetic analysis. We determined the phylogeny of these variants with respect to other reported *A. phagocytophilum* strains (see the supplemental material). For the *ankA* gene, there was 90% bootstrap support for the clade representing the dog variants and California strains being different from the clade representing the New York strains. It does confirm that *ankA* remains a phylogenetically important gene to describe genetic heterogeneity in *A. phagocytophilum*. In addition, there was 90% bootstrap support upholding clade differences between the United States and the Eurasia clades. Interestingly, strains from Germany fell in three different clades, as was also shown by von Loewenich et al. (31). For

TABLE 2. Unique sets of polymorphisms in the *gltA* genes associated with *A. phagocytophilum* variants

GenBank no.[a] or set	Nucleotide at position:					
	66	267	269	314	598	696
AF304136	G	G	G	T	G	G
Alpha	G	T	G	C	G	G
Beta	A	T	A	C	A	A

[a] GenBank accession number used to identify nucleotide positions.

TABLE 4. Combination of multiple gene polymorphisms in different dog samples

Sample	16S rRNA variant(s)[a]	*ankA* set	*gltA* set	*groESL* set(s)
2	v1, v2, v4	Alpha	Alpha	Alpha
3	v1, v4	Alpha	Alpha	Alpha
4	v4	Alpha	Alpha	Alpha, beta
5	v3	Beta	Beta	Alpha or beta
6	v2, v4	Alpha	Alpha	Alpha, gamma
7	v4	Alpha	Alpha	Alpha
8	v4, v5	Alpha	Alpha	Alpha

[a] 16S rRNA based on genetic variants have been described by Poitout et al. (22).

groESL, the gamma set of polymorphisms made a separate clade and clustered with strains from Sardinia (1) but were different from the alpha and beta sets. Otherwise, clade structures were similar to that of the *ankA* gene. The *gltA* tree suggested that variants representing the beta set were closer to strains from California than the New York and Minnesota strains. In summary, Washington variants clustered with U.S. strains in general, except for the *groESL* gamma set, which clustered outside the U.S. strains (see the supplemental material). In general, U.S. strains formed a separate clade from the China clade and the European clade for all genes examined. The data suggest a limited phylogenetic variation within the U.S. clades to distinguish the strains when each gene is examined separately. There is, however, adequate genetic signal, which suggests some differences among the variants but which is more obvious in clades representing European and Chinese strains and possibly the Russian strains. It needs to be pointed out than any phylogenetic relationships based on the gene sequence should be interpreted with caution since recombination is extant among bacterial species (11). In this case, phylogenetic analysis was based on three noncontiguous genes located at significant distances on the *A. phagocytophilum* genome (12). The agreement among the three genes suggests that no recombination event has confounded the analysis.

The strength of our approach to distinguish genotypes from different strains relies on the buildup of a large sequence database of 16S rRNA, *ankA*, *gltA*, and *groESL* genes. Therefore, we recommend that investigators try to obtain sequence data for these genes of *A. phagocytophilum* whenever feasible.

Little is known about the population structure of *A. phagocytophilum* with regard to host tropisms, virulence profiles, and susceptibility to different antibiotics. This limitation is mostly due to approaches available to work on an obligate intracellular pathogen such as *A. phagocytophilum*. It has been difficult to identify phenotypic markers for strain discrimination for this pathogen. Therefore, genotypic based differences could be exploited to address some of the host-pathogen specificity issues. Previous studies have attempted to exploit the polymorphisms in the 16S rRNA gene as a marker to distinguish the virulent and variant *A. phagocytophilum* strains (19, 22, 27). Besides the 16S rRNA gene, *ankA*, *gltA*, and *groESL* are among other polymorphic genes that have been exploited independently to investigate phylogenetic relatedness for *A. phagocytophilum* (8, 9, 13–15, 18, 28, 31). Our multigene analysis approach with the use of known variants provides an opportunity to more narrowly define *A. phagocytophilum* genotypic differences at the population genetics level. This approach will help in our understanding of the clonal concept for this emerging pathogen.

We acknowledge Marshfield Clinic Research Foundation for financial support.

We thank Linda Weis and Alice Stargardt for assistance in preparation of the manuscript.

REFERENCES

1. **Alberti, A., R. Zobba, B. Chessa, M. F. Addis, O. Sparagano, M. L. Pinna Parpaglia, T. Cubeddu, G. Pintori, and M. Pittau.** 2005. Equine and canine *Anaplasma phagocytophilum* strains isolated on the island of Sardinia (Italy) are phylogenetically related to pathogenic strains from the United States. Appl. Environ. Microbiol. **71:**6418–6422.
2. **Asanovich, K. M., J. S. Bakken, J. E. Madigan, M. Aguero-Rosenfeld, G. P. Wormser, and J. S. Dumler.** 1997. Antigenic diversity of granulocytic *Ehrlichia* isolates from humans in Wisconsin and New York and a horse in California. J. Infect. Dis. **176:**1029–1034.
3. **Barlough, J. E., J. E. Madigan, D. R. Turoff, J. R. Clover, S. M. Shelly, and J. S. Dumler.** 1997. An *Ehrlichia* strain from a llama (*Lama glama*) and llama-associated ticks (*Ixodes pacificus*). J. Clin. Microbiol. **35:**1005–1007.
4. **Belongia, E. A., K. D. Reed, P. D. Mitchell, C. P. Kolbert, D. H. Persing, J. S. Gill, and J. J. Kazmierczak.** 1997. Prevalence of granulocytic *Ehrlichia* infection among white-tailed deer in Wisconsin. J. Clin. Microbiol. **35:**1465–1468.
5. **Björsdorff, A., L. Svendenius, J. H. Owens, and R. F. Massung.** 1999. Feline granulocytic ehrlichiosis—a report of a new clinical entity and characterisation of the infectious agent. J. Small Anim. Pract. **40:**20–24.
6. **Björsdorff, A., J. Berglund, B. E. Kristiansen, C. Soderstrom, and I. Eliasson.** 1999. Varying clinical picture and course of human granulocytic ehrlichiosis. Twelve Scandinavian cases of the new tick-borne zoonosis are presented. Lakartidningen **96:**4200–4204.
7. **Björsdorff, A., B. Bagert, R. F. Massung, A. Gusa, and I. Eliasson.** 2002. Isolation and characterization of two European strains of *Ehrlichia phagocytophila* of equine origin. Clin. Diagn. Lab. Immunol. **9:**341–343.
8. **Caturegli, P., K. M. Asanovich, J. J. Walls, J. S. Bakken, J. E. Madigan, V. L. Popov, and J. S. Dumler.** 2000. *ankA*: an *Ehrlichia phagocytophila* group gene encoding a cytoplasmic protein antigen with ankyrin repeats. Infect. Immun. **68:**5277–5283.
9. **Chae, J. S., J. E. Foley, J. S. Dumler, and J. E. Madigan.** 2000. Comparison of the nucleotide sequences of 16S rRNA, 444 *Ep-ank*, and *groESL* heat shock operon genes in naturally occurring *Ehrlichia equi* and human granulocytic ehrlichiosis agent isolates from northern California. J. Clin. Microbiol. **38:**1364–1369.
10. **Foley, J. E., L. Crawford-Miksza, J. S. Dumler, C. Glaser, J. S Chae, E. Yeh, D. Schnurr, R. Hood, W. Hunter, and J. E. Madigan.** 1999. Human granulocytic ehrlichiosis in northern California: two case descriptions with genetic analysis of the *Ehrlichiae*. Clin. Infect. Dis. **29:**388–392.
11. **Hall, B., and M. Barlow.** 2006. Phylogenetic analysis as a tool in molecular epidemiology of infectious diseases. Ann. Epidemiol. **16:**157–169.
12. **Hotopp, J. C., M. Lin, R. Madupu, J. Crabtree, S. V. Angiuoli, J. Eisen, R. Seshadri, Q. Ren, M. Wu, T. R. Utterback, S. Smith, M. Lewis, H. Khouri, C. Zhang, H. Niu, Q. Lin, N. Ohashi, N. Zhi, W. Nelson, L. M. Brinkac, R. J. Dodson, M. J. Rosovitz, J. Sundaram, S. C. Daugherty, T. Davidsen, A. S. Durkin, M. Gwinn, D. H. Haft, J. D. Selengut, S. A. Sullivan, N. Zafar, L. Zhou, F. Benahmed, H. Forberger, R. Halpin, S. Mulligan, J. Robinson, O. White, Y. Rikihisa, and H. Tettelin.** 2006. Comparative genomics of emerging human ehrlichiosis agents. PLoS Genet. **2:**e21.
13. **Inokuma, H., P. Brouqui, M. Drancourt, and D. Raoult.** 2001. Citrate synthase gene sequence: a new tool for phylogenetic analysis and identification of *Ehrlichia*. J. Clin. Microbiol. **39:**3031–3039.
14. **Inokuma, H., K. Fujii, M. Okuda, T. Onishi, J.-P. Beaufils, D. Raoult, and P. Brouqui.** 2002. Determination of the nucleotide sequences of heat shock operon *groESL* and the citrate synthase gene (*gltA*) of *Anaplasma* (*Ehrlichia*) *platys* for phylogenetic and diagnostic studies. Clin. Diagn. Lab. Immunol. **9:**1132–1136.
15. **Lee, J. H., H. S. Park, W. J. Jang, S. E. Koh, J. M. Kim, S. K. Shim, M. Y. Park, Y. W. Kim, B. J. Kim, Y. H. Kook, K. H. Park, and S. H. Lee.** 2003. Differentiation of rickettsiae by *groEL* gene analysis. J. Clin. Microbiol. **41:**2952–2960.
16. **Lotric-Furlan, S., T. Avsic-Zupanc, M. Petrovec, W. L. Nicholson, J. W. Sumner, J. E. Childs, and F. Strle.** 2001. Clinical and serological follow-up of patients with human granulocytic ehrlichiosis in Slovenia. Clin. Diagn. Lab. Immunol. **8:**899–903.
17. **Maddison, D. R., and W. P. Maddison.** 2001. MacClade 4: analysis of phylogeny and character evolution, version 4.02. Sinauer Associates, Sunderland, MA.
18. **Massung, R. F., J. H. Owens, D. Ross, K. D. Reed, M. Petrovec, A. Björsdorff, R. T. Coughlin, G. A. Beltz, and C. I. Murphy.** 2000. Sequence analysis of the *ank* gene of granulocytic ehrlichiae. J. Clin. Microbiol. **38:**2917–2922.
19. **Massung, R. F., M. J. Mauel, J. H. Owens, N. Allan, J. W. Courtney, K. C. Stafford III, and T. N. Mather.** 2002. Genetic variants of *Ehrlichia phagocytophila*, Rhode Island and Connecticut. Emerg. Infect. Dis. **8:**467–472.
20. **Massung, R. F., R. A. Priestley, N. J. Miller, T. N. Mather, and M. L. Levin.** 2003. Inability of a variant strain of *Anaplasma phagocytophilum* to infect mice. J. Infect. Dis. **188:**1757–1763.
21. **Massung, R. F., T. N. Mather, and M. L. Levin.** 2006. Reservoir competency of goats for the Ap-variant 1 strain of *Anaplasma phagocytophilum*. Infect. Immun. **74:**1373–1375.
22. **Poitout, F. M., J. K. Shinozaki, P. J. Stockwell, C. J. Holland, and S. K. Shukla.** 2005. Genetic variants of *Anaplasma phagocytophilum* infecting dogs in western Washington State. J. Clin. Microbiol. **43:**796–801.
23. **Pusterla, N., C. C. Chang, B. B. Chomel, J. S. Chae, J. E. Foley, E. DeRock, V. L. Kramer, H. Lutz, and J. E. Madigan.** 2000. Serologic and molecular evidence of *Ehrlichia* spp. in coyotes in California. J. Wildl. Dis. **36:**494–499.
24. **Relman, D. A.** 1993. Universal bacterial 16S rDNA amplification and sequencing, p. 489–495. *In* D. H. Persing, T. F. Smith, F. C. Tenover, and T. J.

White (ed.), Diagnostic molecular microbiology. Principles and applications. American Society for Microbiology, Washington, DC.
25. **Remy, V., Y. Hansmann, S. De Martino, D. Christmann, and P. Brouqui.** 2003. Human anaplasmosis presenting as atypical pneumonitis in France. Clin. Infect. Dis. **37:**846–848.
26. **Roux, V., E. Rydkina, M. Eremeeva, and D. Raoult.** 1997. Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the rickettsiae. Int. J. Syst. Bacteriol. **47:**252–261.
27. **Stuen, S., I. Van De Pol, K. Bergstrom, and L. M. Schouls.** 2002. Identification of *Anaplasma phagocytophila* (formerly *Ehrlichia phagocytophila*) variants in blood from sheep in Norway. J. Clin. Microbiol. **40:**3192–3197.
28. **Sumner, J. W., W. L. Nicholson, and R. F. Massung.** 1997. PCR amplification and comparison of nucleotide sequences from the *groESL* heat shock operon of *Ehrlichia* species. J. Clin. Microbiol. **35:**2087–2092.
29. **Sumner, J. W., G. A. Storch, R. S. Buller, A. M. Liddell, S. L. Stockham, Y. Rikihisa, S. Messenger, and C. D. Paddock.** 2000. PCR amplification and phylogenetic analysis of *groESL* operon sequences from *Ehrlichia ewingii* and *Ehrlichia muris*. J. Clin. Microbiol. **38:**2746–2749.
30. **Swofford, D. L.** 2003. PAUP*: phylogenetic analysis using parsimony (*and other methods), version 4. Sinauer Associates, Sunderland, MA.
31. **von Loewenich, F. D., B. U. Baumgarten, K. Schroppel, W. Geissdorfer, M. Rollinghoff, and C. Bogdan.** 2003. High diversity of *ankA* sequences of *Anaplasma phagocytophilum* among *Ixodes ricinus* ticks in Germany. J. Clin. Microbiol. **41:**5033–5040.
32. **Walls, J. J., P. Caturegli, J. S. Bakken, K. M. Asanovich, and J. S. Dumler.** 2000. Improved sensitivity of PCR for diagnosis of human granulocytic ehrlichiosis using *epank1* genes of *Ehrlichia phagocytophila*-group ehrlichiae. J. Clin. Microbiol. **38:**354–356.
33. **Weber, R., N. Pusterla, M. Loy, C. M. Leutenegger, G. Schar, D. Baumann, C. Wolfensberger, and H. Lutz.** 2000. Serologic and clinical evidence for endemic occurrences of human granulocytic ehrlichiosis in north-eastern Switzerland. Schweiz. Med. Wochenschr. **130:**1462–1470.