6076–6091 *Nucleic Acids Research, 2009, Vol. 37, No. 18* *Published online 6 August 2009*
doi:10.1093/nar/gkp642

Analysis of complete genome sequence of *Neorickettsia risticii*: causative agent of Potomac horse fever

Mingqun Lin, Chunbin Zhang, Kathryn Gibson and Yasuko Rikihisa*

Department of Veterinary Biosciences, The Ohio State University, Columbus, OH 43210, USA

Received May 1, 2009; Revised July 17, 2009; Accepted July 19, 2009

ABSTRACT

Neorickettsia risticii **is an obligate intracellular bacterium of the trematodes and mammals. Horses develop Potomac horse fever (PHF) when they ingest aquatic insects containing encysted *N. risticii*-infected trematodes. The complete genome sequence of *N. risticii* Illinois consists of a single circular chromosome of 879 977 bp and encodes 38 RNA species and 898 proteins. Although *N. risticii* has limited ability to synthesize amino acids and lacks many metabolic pathways, it is capable of making major vitamins, cofactors and nucleotides. Comparison with its closely related human pathogen *N. sennetsu* showed that 758 (88.2%) of protein-coding genes are conserved between *N. risticii* and *N. sennetsu*. Four-way comparison of genes among *N. risticii* and other Anaplasmataceae showed that most genes are either shared among Anaplasmataceae (525 orthologs that generally associated with housekeeping functions), or specific to each genome (>200 genes that are mostly hypothetical proteins). Genes potentially involved in the pathogenesis of *N. risticii* were identified, including those encoding putative outer membrane proteins, two-component systems and a type IV secretion system (T4SS). The bipolar localization of T4SS pilus protein VirB2 on the bacterial surface was demonstrated for the first time in obligate intracellular bacteria. These data provide insights toward genomic potential of *N. risticii* and intracellular parasitism, and facilitate our understanding of PHF pathogenesis.**

INTRODUCTION

Characterized by fever, depression, anorexia, dehydration, watery diarrhea, laminitis and/or abortion, Potomac horse fever (PHF) is an acute, often severe to fatal systemic disease of horses and typically occurs in the warm weather months of middle to late summer (1,2). The outbreak of PHF in the 1970s along the Potomac River in Maryland and Virginia helped to recognize PHF as a new disease entity (3). Subsequent investigations led to the ultrastructural observation of intracellular bacteria similar to *Neorickettsia sennetsu*, the agent of human Sennetsu neorickettsiosis, in intestinal tissues of horses with acute PHF (2,4) and the discovery of a new bacterium, *Neorickettsia risticii* (formerly *Ehrlichia risticii*). *N. risticii* was demonstrated as the causative agent of PHF by fulfilling Koch's postulates (1,5). Currently, PHF is frequently found throughout North America and increasingly recognized in South America, including Brazil and Uruguay (6,7).

In the natural environment, *Neorickettsia* spp. reside inside trematodes, which can be transstadially transmitted through all developmental stages of trematodes and transovarially passed through generations of trematodes. The relationship of *N. risticii* with its trematode host seems to be either commensal or mutualistic, as reproduction of trematodes does not appear to be adversely affected by infection (8). Mammalian infection by *Neorickettsia* spp. occurs by horizontal transmission of the bacterium from trematodes to susceptible mammalian hosts, mostly through ingestion of this bacterium in the metacercarial stage of trematodes encysting in insects or fish (6). In the eastern United States, *N. risticii* is maintained by transstadial and transovarial passage in the digenetic trematode, *Acanthatrium oregonense*, which has a complex life cycle consisting of miracidia and sporocysts in snail hosts (*Elimia virginica*), free-swimming cercariae, metacercariae in aquatic insects (caddisflies, mayflies), and adults that lay eggs in the intestinal lumen of insectivorous bats (6,8–11). Upon ingestion of *N. risticii* in the metacercarial stage of the trematodes in aquatic insects by horses, *N. risticii* is horizontally transmitted from the trematodes to horses and replicates within inclusion bodies inside monocytes, macrophages, mast cells and intestinal epithelial cells (2,11–13). Currently, the only effective treatment of PHF is the administration of

*To whom correspondence should be addressed. Tel: +1 614 292 5661; Fax: +1 614 292 6473; Email: rikihisa.1@osu.edu

Table 1. Biological characteristics of the selected members of the family Anaplasmataceae

Species	Vertebrate host	Invertebrate vector/host[a]	*In vivo*-infected mammalian cells	Diseases	Geographical distribution
N. risticii	Horse, Bat	Digenetic trematode *Acanthatrium oregonense* of snails and aquatic insects (TOP)	Monocytes, macrophages, intestinal epithelial cells, and mast cells	Potomac horse fever	USA, Canada, Brazil, Uruguay
N. sennetsu	Human	Unknown trematodes of snails and fish (TOP, not proven)	Monocytes and macrophages	Sennetsu neorickettsiosis	Southeast Asia
N. helminthoeca	Canidae	Trematode *Nanophyetus salmincola* of snails and fish (TOP)	Monocytes and macrophages	Salmon poisoning disease	California, Washington, Oregon, Idaho, Canada
A. phagocytophilum	Human, Deer, Cat, Rodent, Sheep, Cattle, Horse, Dog, Wild boar, Llama	*Ixodes* Ticks (*I. scapularis, I. ricinus*, *I. persulcatus*)	Granulocytes and endothelial cells	Human granulocytic anaplasmosis	USA, Europe, Asia
E. chaffeensis	Human, Deer, Dog	Ticks (*Amblyomma americanum*)	Monocytes and macrophages	Human monocytic ehrlichiosis	USA, Israel, Europe, Africa, South and Central America
The *Wolbachia* endosymbiont of *Brugia malayi*	N/A	Filarial nematode in insects (TOP)	N/A	(River blindness/ inflammation)[b]	World Wide

[a]Transmission mode: Except for *Anaplasma* and *Ehrlichia* species, all these listed organisms can be transovarially transmitted (TOP, Transovarial passage).
[b]Filarial nematode causes debilitating inflammatory diseases such as river blindness and lymphatic filariasis. However, the predominant inflammatory response in the cornea is due to a species of endosymbiotic *Wolbachia* bacteria (*Science* 295: 1982, 2002).

broad-spectrum tetracycline antibiotics in the early stages of the disease (11). Although a vaccine against PHF has been marketed, PHF continues to cause widespread infections, probably due to both the insufficient immunity developed by the vaccination and the antigenic variation of *N. risticii* strains in the field (14,15).

So far, *Neorickettsia* spp. that cause significant illness in mammals have been studied sufficiently to be officially classified. These are *N. risticii*, *N. sennetsu* (formerly *Ehrlichia sennetsu* and *Rickettsia sennetsu*) and *N. helminthoeca*, the agent of Salmon poisoning disease in dogs (Table 1) (6,16,17). The 16S rRNA-based phylogenic tree shows that *Neorickettsia* species reside in a clade separated from other Anaplasmataceae in the order Rickettsiales, including *Ehrlichia chaffeensis*, *Anaplasma phagocytophilum* and the *Wolbachia* endosymbiont of *Brugia malayi* (wBm) (Figure 1). While *N. risticii* infects a trematode that uses an aquatic insect as an intermediate host in North America, *N. sennetsu* infects a trematode that likely uses a fish as an intermediate host in Southeast Asia (6,11). Despite distinct trematode and mammalian hosts, pathogenesis and geographic ranges, phylogenetic analysis based on the 16S rRNA sequence indicates that there is only 0.7% divergence between *N. risticii* and *N. sennetsu* (Figure 1).

In the present study, the *N. risticii* genome is sequenced, compared with those of other members of the family Anaplasmataceae, especially *N. sennetsu*: the only sequenced member of the *Neorickettsia* genus with unknown trematode association, and potential virulence factors and novel outer membrane proteins are identified. While *N. risticii* is the newest member of the genus *Neorickettsia*, of all *Neorickettsia* spp., it has the broadest geographic distribution, inflicting the greatest economical and emotional loss, and the best information available for pathogenesis and immune responses (6). These genome sequence data will be critical for enhancing our knowledge of this obligate intracellular bacterium, providing tools for better understanding PHF pathogenesis, and the development of effective vaccines.

MATERIALS AND METHODS

Culture and purification of *N. risticii* genomic DNA

Neorickettsia risticii Illinois^T was obtained from the Naval Medical Research Center (Bethesda, MD) (11). *Neorickettsia risticii* Illinois was propagated in P388D$_1$ cells, a murine monocytic leukemia cell line, in RPMI 1640 medium supplemented with 10% fetal bovine serum and 2 mM L-glutamine in a humidified 5% CO_2–95% air atmosphere at 37°C as previously described (18). Bacterial cells were liberated from the infected host cells using Dounce homogenization, and purified by differential centrifugation and Percoll density gradient centrifugation (17). Any specimens with host nuclei contamination were excluded. From these isolated bacteria, phenol extraction was used to purify DNA that was minimally fragmented and free of host-cell DNA. Levels of host DNA contamination were verified to be less than 0.01% by PCR using host G3PDH-specific primers.

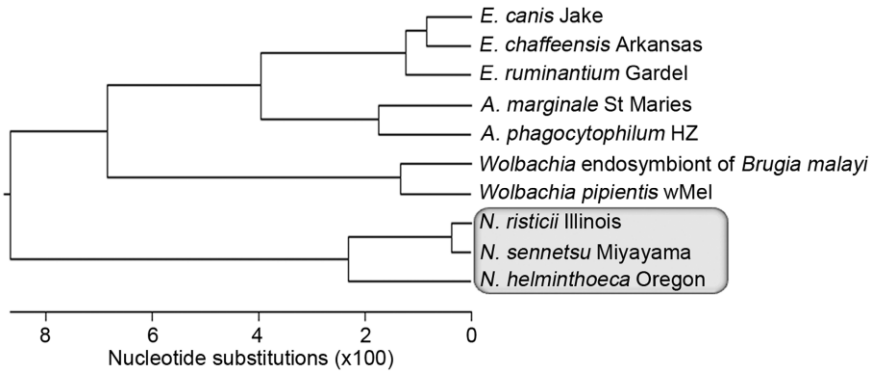

Figure 1. Phylogenetic tree of the family Anaplasmataceae. 16S rRNA sequences of members of the family Anaplasmataceae were aligned using the Clustal W method, and a phylogenetic tree was built. Gray box highlights *Neorickettsia* species.

Genome sequencing, gap closure and assembly

The genome of *N. risticii* Illinois was sequenced with shotgun sequencing at 20× coverage using a pyrosequencing protocol in a microfabricated high-density picoliter reaction (454 Life Sciences, Branford, CT). Eight major contigs (>16 kb each) and five minor contigs (629–5105 bp) were obtained, and gaps were closed by primer walking. All PCR primers were designed around 200 bp from the ends of each contigs, with the positions of contigs in the *N. risticii* genome predicted by blasting those to the *N. sennetsu* genome (GenBank No. NC_007798). The whole genome was assembled from the contigs and the sequenced gap segments by the SeqMan program from the Lasergene DNAStar software package (Madison, WI). The GC-skew was calculated as (C–G)/(C + G) in windows of 1000 bp along the chromosome (19).

Genome annotation and comparison with other related organisms

The DNA sequence was submitted to the JCVI (J. Craig Venter Institute) Annotation Service, where it was run through JCVI's prokaryotic annotation pipeline (http://www.jcvi.org/cms/research/projects/annotation-service/). Included in the pipeline is gene finding with Glimmer 2, Blast-extend-repraze (BER) searches, HMM (a hidden Markov model) searches, TMHMM (transmembrane helix prediction) searches, SignalP predictions and automatic annotations from AutoAnnotate. Potential protein and RNA-coding sequences were predicted and annotated from these tools, including feature identification (e.g. protein motifs), and assignment of database matches and functional role categories to genes (17). The manual annotation tool Manatee was downloaded from SourceForge (http://manatee.sourceforge.net) and used to manually review and curate the output from the prokaryotic pipeline of the JCVI Annotation Service. Ribosome binding sites (RBS) were determined by RBSFinder (ftp://ftp.tigr.org/pub/software/RBSfinder/). The K_a/K_s ratio was determined using WSPMaker with the sliding window method (http://wspmaker.kobic.kr/), a web based tool to calculate and display the selection pressure in sub-regions of two orthologous protein-coding DNA sequences (20).

Phylogenetic trees were constructed based on sequence alignment by the Clustal W method using the MegAlign program from the Lasergene package. Comparative analysis of *N. risticii* and related organisms (*N. sennetsu*, *E. chaffeensis*, *A. phagocytophilum* and *Wolbachia* wBm) was performed by using reciprocal BLASTP. *N. risticii* genes with or without homology to other related organisms were identified by reciprocal BLASTP hits with cutoff scores of $E < 10^{-5}$. Signal peptides and transmembrane helices were predicted by using SignalP 3.0 (21) and TMHMM 2.0 (22) set at default values.

Analysis of VirB2 expression in *N. risticii*-infected $P388D_1$ cells

To detect gene expression of *virB2* in *N. risticii*, total RNA was extracted from *N. risticii*-infected $P388D_1$ cells at 3 days post-infection (p.i.) and RT–PCR was carried out using specific primers spanning *virB2.1* and *virB2.2* genes (Supplementary Table 1) as described previously (23). The expression of VirB2 protein was confirmed by western blot analysis as described previously (23), using custom-raised rabbit antisera against 15-mer peptide ($aa^{32–46}$: AGPDKDDSIVSRVIC) of *N. risticii* VirB2-1 (NRI_0738) (Sigma-Genosys, Woodlands, TX).

Double immunofluorescence labeling

Neorickettsia risticii-infected $P388D_1$ cells at 2 days p.i., or *N. risticii* organisms purified as previously described (24), were centrifuged onto glass slides using a Shandon Cytospin 4 cytocentrifuge (Thermo Fisher Scientific, Kalamazoo, MI) and fixed in 4% paraformaldehyde at room temperature for 15 min. *Neorickettsia risticii*-infected cells were permeabilized with 0.1% saponin for 15 min and washed, whereas purified *N. risticii* organisms were subjected to labeling without saponin permeabilization. Samples were incubated with horse anti-*N. risticii* serum and rabbit anti-*N. risticii* VirB2 (NRI_0738) antibody in phosphate-buffered saline (PBS) containing 1% bovine serum albumin for 1 h at room temperature. After washing with PBS, cells were labeled with FITC-conjugated goat anti-horse (Jackson

ImmunoResearch, West Grove, PA) and Alexa Fluor 555–conjugated goat anti-rabbit (Invitrogen, Carlsbad, CA) secondary antibodies for 1 h. Fluorescence images were analyzed and captured by a SPOT CCD digital camera (Diagnostic Instruments, Sterling Heights, MI) connected to a Nikon Eclipse E400 fluorescent microscope with a xenon-mercury light source (Nikon Instruments Inc, Melville, NY).

GenBank accession number

The sequence data described here have been deposited in GenBank (accession number CP001431).

RESULTS AND DISCUSSION

General features of the genome

The genome of *N. risticii* consists of a single circular chromosome spanning 879 977 nt and has a G + C content of 41.3% (Table 2). The replication origin of the *N. risticii* was predicted based on one of the GC-skew shift points and the location of *parB* (Figure 2). This putative origin coincides with those of closely related species *E. chaffeensis* and *N. sennetsu* (17). The *N. risticii* and *N. sennetsu* genomes are smaller than those of other members in the family Anaplasmataceae, which are ~1.0–1.5 Mb. In agreement with small 16S rRNA gene sequence divergence between *N. risticii* and *N. sennetsu* (Figure 1), the synteny plot between *N. risticii* and *N. sennetsu* indicates that these two genomes exhibit almost complete synteny, whereas no significant synteny is present with other members of the Anaplasmataceae (Figure 3).

The *N. risticii* genome encodes one copy of 5S, 16S and 23S rRNA genes. Similar to the other Anaplasmataceae sequenced, the 5S and 23S rRNA genes form an operon (17), while the 16S rRNA gene is separated by ~0.7 Mb and followed by *sdh*CD, as in *A. phagocytophilum* (25). Thirty-three tRNA genes are identified, which include cognates for all 20 amino acids. *N. risticii* has a smaller number of predicted genes and a lower coding density (86.8%) than *N. sennetsu* (Table 2). Of 898 predicted protein-coding open reading frames (ORFs), 554 genes are assigned with probable functions based on similarity searches. Nearly 40% of the predicted ORFs (345 genes) in the genome are annotated as 'hypothetical' genes, either with conserved domains or unknown functions (Table 2 and Figure 2, gray bars in first two circles).

Analysis of the *N. risticii* genome identified thirteen pairs of overlaps in protein-coding ORFs, and seven pairs of overlaps in RNA- and protein-coding genes (Supplementary Figure 1). Although most of the protein-coding ORFs that overlap with RNA genes encode small proteins (<50 aa) without any known functions, one encodes replicative DNA helicase (*dnaB*, NRI_0529), which potentially has essential functions for bacterial replication, and overlaps with an essential tRNA gene that decodes Proline (Supplementary Figure 1B). Overlapping genes are detected primarily in parasitic or symbiotic bacteria and are believed to be a consequence of the reduction of originally larger genomes (26,27). Whether these overlapping genes are expressed by *N. risticii* remains to be experimentally verified.

The *N. risticii* genome has 110 ORFs that encode small hypothetical proteins with fewer than 60 amino acids. By RBSFinder analysis, 43 out of 110 (39%) do not have RBS upstream of the start codon. It is possible that translation of these leaderless mRNAs can be initiated via a novel pathway that involves the preassembled 70S ribosome rather than via a free 30S subunit (28). To determine whether these small protein-coding ORFs have evolved as typical coding genes with purifying selection restricting nonsynonymous changes, the K_a/K_s ratio was determined using WSPMaker with the sliding window method (20). The K_a/K_s ratio represents the rate of nonsynonymous substitutions (K_a) to that of synonymous substitutions (K_s), which can be used as an indicator of selective pressure acting on a protein-coding gene (29). Since WSPMaker takes orthologous protein-coding sequence pairs as input, we first used BLASTP to determine the small hypothetical proteins that are conserved between *N. risticii* and *N. sennetsu*. The resulting 26 orthologous ORF pairs were then concatenated with the stop codons removed (Supplementary Figure 2). Results demonstrated that,

Table 2. Genome properties in the selected members of the family Anaplasmataceae

Organisms	NRI	NSE	APH	ECH	WBM
ORFs	898	935	1369	1115	1218
tRNA	33	33	37	37	34
rRNA	3	3	3	3	3
sRNA	2	2	2	3	1
Size	879 977	859 006	1 471 282	1 176 248	1 080 084
GC (%)	41.3	41.1	41.6	30.1	34.2
Average gene length	841	804	775	840	676
Coding (%)[a]	86.9	89.2	90.5	87.6	77.4
Assigned functions	554	534	756	608	664
Unknown functions	344	401	612	506	554

APH, *Anaplasma phagocytophilum*; ECH, *Ehrlichia chaffeensis*; NSE, *Neorickettsia sennetsu*; NRI, *Neorickettsia risticii*; WBM, the *Wolbachia* endosymbiont of *Brugia malayi*.
[a]Percent coding includes tRNA, rRNA, sRNA and all protein-coding genes.

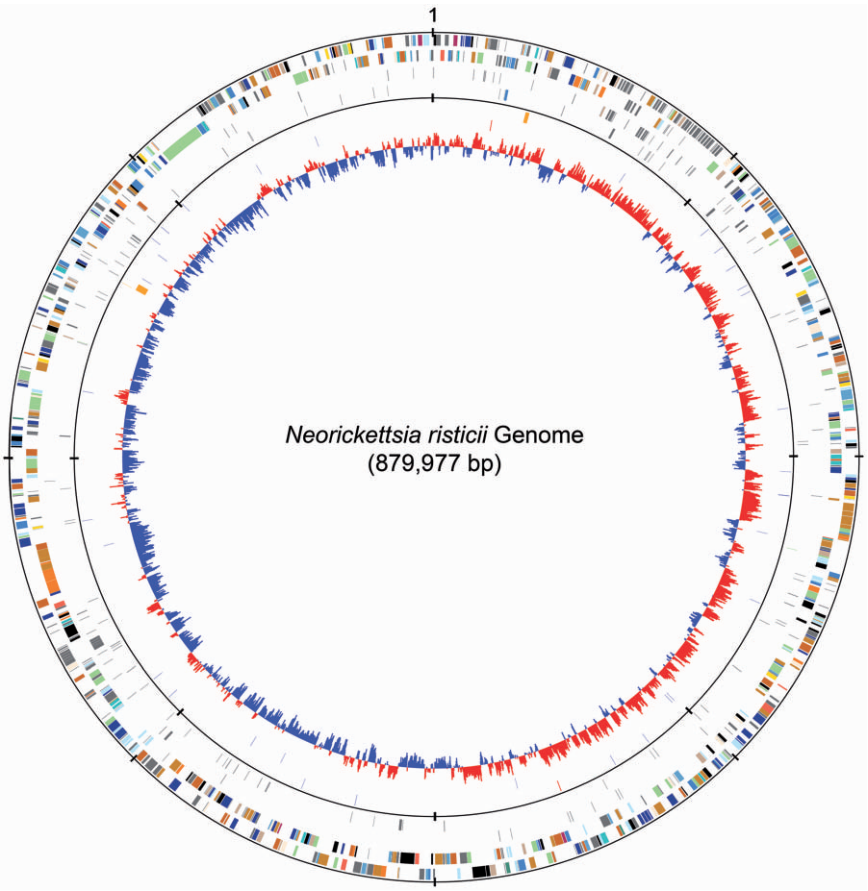

Figure 2. Circular representation of the genome of *N. risticii*. From outside to inside, the first two circles represent predicted protein-coding sequences (ORFs) on the plus and minus strands, respectively. Colors indicate the role categories of ORFs: dark gray: hypothetical proteins or proteins with unknown functions; gold: amino-acid and protein biosynthesis; sky blue: purines, pyrimidines, nucleosides and nucleotides; cyan: fatty acid and phospholipid metabolism; light blue: biosynthesis of cofactors, prosthetic groups and carriers; aquamarine: central intermediary metabolism; royal blue: energy metabolism; pink: transport and binding proteins; dark orange: DNA metabolism and transcription; pale green: protein fate; tomato: regulatory functions and signal transduction; peach puff: cell envelope; pink: cellular processes; maroon: mobile and extrachromosomal element functions. The third and fourth circles show unique ORFs compared to *N. sennetsu*. The fifth and sixth circles represent RNA genes, including tRNAs (blue), rRNAs (orange), and sRNAs (red). The seventh circle represents G–C skew values [(G − C)/(G + C)] with a windows size of 1 kb.

although the average K_a/K_s ratio of the concatenated 26 orthologous ORF pairs of *N. risticii* and *N. sennetsu* is 0.71, four regions, corresponding to six *N. risticii* protein-coding ORFs, exhibit K_a/K_s value >1 (Supplementary Figure 2). It suggests that these ORFs may have positive selection for advantageous mutations.

It has been suggested that around 10–30% of small protein-coding ORFs (<300 bp) do not actually encode proteins (30,31). As an obligate intracellular bacterium with a genome of <900 kb, *N. risticii* may share genomic features with closely related members in the family Anaplasmataceae. A whole genome transcription profiling study of *A. phagocytophilum* showed that ~70% of ORFs were significantly transcribed, including 342 out of 409 ORFs (84%) of fewer than 180 bp (32). Furthermore, the global proteomic analysis of *E. chaffeensis* showed that more than 90% of 1115 predicted proteins were expressed, including 130 (62%) out of 209 proteins of fewer than 60 aa, most of which are hypothetical proteins (33). These data suggest that most of *N. risticii* small ORFs could also be transcribed similar to *E. chaffeensis* and *A. phagocytophilum*. As there is no definitive way to determine functional ORFs other than the experimental analysis, and all future analyses such as microarray and proteomics depend on the completeness of ORFs in curated genome sequences, we opt to retain all predicted ORFs in our annotation to facilitate future studies.

Comparison of protein-coding genes among *N. risticii* and other Anaplasmataceae

In order to compare the genome content of *N. risticii* to representative members of Anaplasmataceae including *N. sennetsu*, *E. chaffeensis*, *A. phagocytophilum* and *Wolbachia* wBm, two-, three- and four-way comparisons were performed, and homologous clusters were constructed. Two-way comparison between *N. risticii* and *N. sennetsu* shows that these two genomes have 758 protein-coding ORFs that are orthologous to each other (88.2% of total protein-coding ORFs) (Figure 4A).

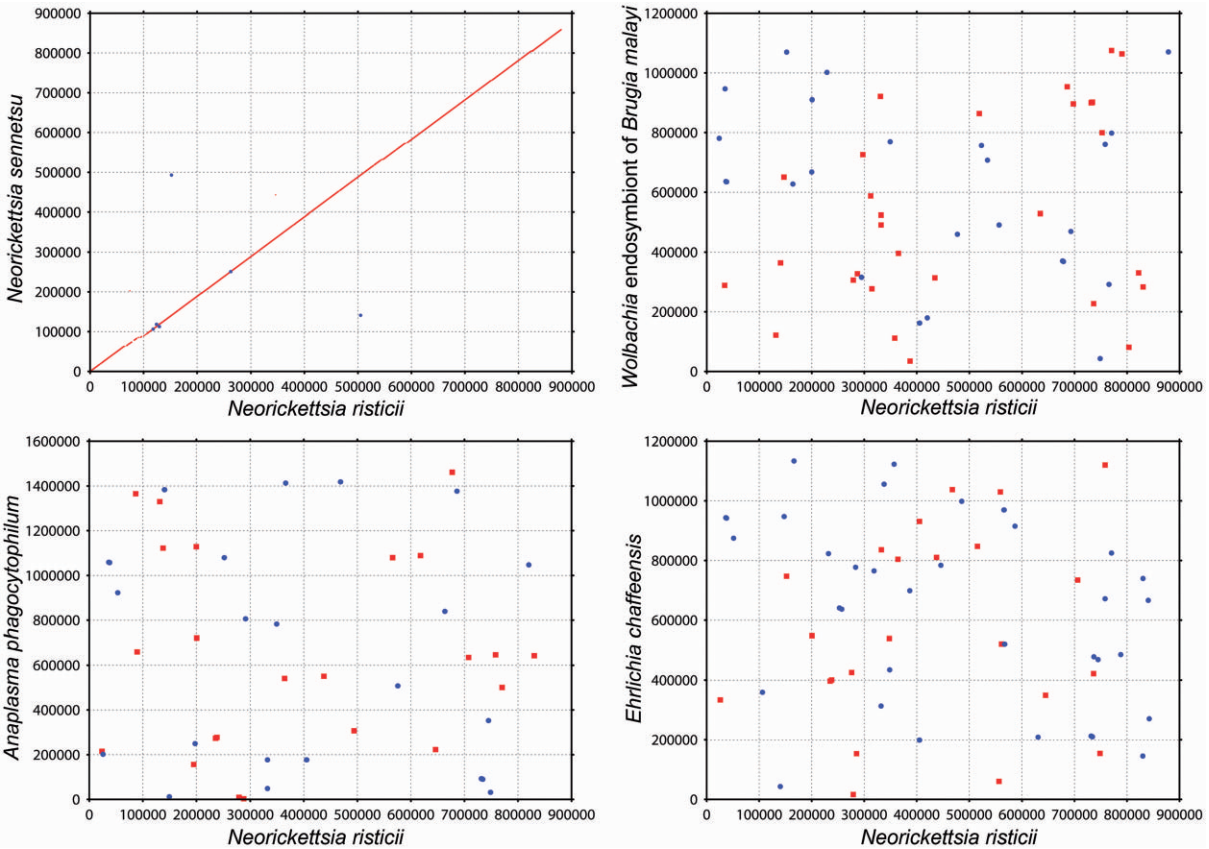

Figure 3. Synteny plots between *N. risticii* Illinois (horizontal axis) and *N. sennetsu* Miyayama, *A. phagocytophilum* HZ, *E. chaffeensis* Arkansas, and the *Wolbachia* endosymbiont of *Brugia malayi*. Numbers represent base pairs. Each dot represents a pair of probable sequence fragments defined as reciprocal BLAST best hits with *E*-value <0.001 (red: sequences match at the forward strand; blue: sequences match at the reverse complemented strand).

Despite almost complete genome synteny between *N. risticii* and *N. sennetsu*, 140 (15.6% of total *N. risticii* ORFs, Figure 2, third and fourth circles) and 177 genes (18.9% of total *N. sennetsu* ORFs) are unique to each organism, respectively. Among them, 120 *N. risticii* genes (13.4%) do not exhibit similarity to any genes of either bacterial or eukaryotic origin (Table 3 and Supplementary Table 2). Functions of these predicted proteins are unknown, and their molecular masses range from 4.1 to 92.3 kDa. Of note, three *N. risticii* genes (0.3% of total genes) match only to genes in other Anaplasmataceae genomes, but not to genes in the *N. sennetsu* genome (Supplementary Table 2). Two ORFs encoding Na^+/H^+ antiporter MnhB subunit-related proteins (NRI_0032/NRI_0033) are found only in members of Anaplasmataceae that primarily infect animals, such as *E. ruminantium*, *E. canis* and *A. marginale*, but not in human pathogens such as *N. sennetsu*, *E. chaffeensis*, or *A. phagocytophilum* human isolate. One conserved hypothetical protein matches to *Wolbachia* and *Ehrlichia* spp., with the characteristic C-terminal basic amino-acid motif for a type IV secretion system (T4SS) substrate (NRI_0703). Interestingly, *N. risticii* is the only sequenced species in the order Rickettsiales that encodes a complete DNA photolyase (deoxyribodipyrimidine photolyase, NRI_0805): an alternative mechanism for repairing UV-induced DNA damage (Supplementary Table 2). Although *N. sennetsu* also encodes a DNA photolyase, it contains a point mutation at nt^{814} (CAA → TAA), which results in a premature stop codon at aa^{271} (full-length protein length: ~470 aa) and renders this ORF nonfunctional.

Four-way comparison of genes among *N. risticii* and selected Anaplasmataceae members shows that most *N. risticii* genes are either shared among all Anaplasmataceae (525 orthologs) or specific to each genome (>200 ORFs) (Figure 4B). Analysis of these genes for functional categories indicates that most of the shared genes are associated with housekeeping functions (Table 3 and Supplementary Table 3). Of *N. risticii*-specific genes detected in this four-way comparison, the vast majority of these genes encode hypothetical, conserved hypothetical, and conserved domain proteins, as well as uncharacterized membrane proteins or lipoproteins (Table 3 and Supplementary Table 4). Two- and three-way comparisons show that most genes in these four genomes are either conserved among *E. chaffeensis*, *A. phagocytophilum*, and *Wolbachia* wBm (76 ortholog clusters), or between *E. chaffeensis* and *A. phagocytophilum* (53 ortholog clusters) (Figure 4B and

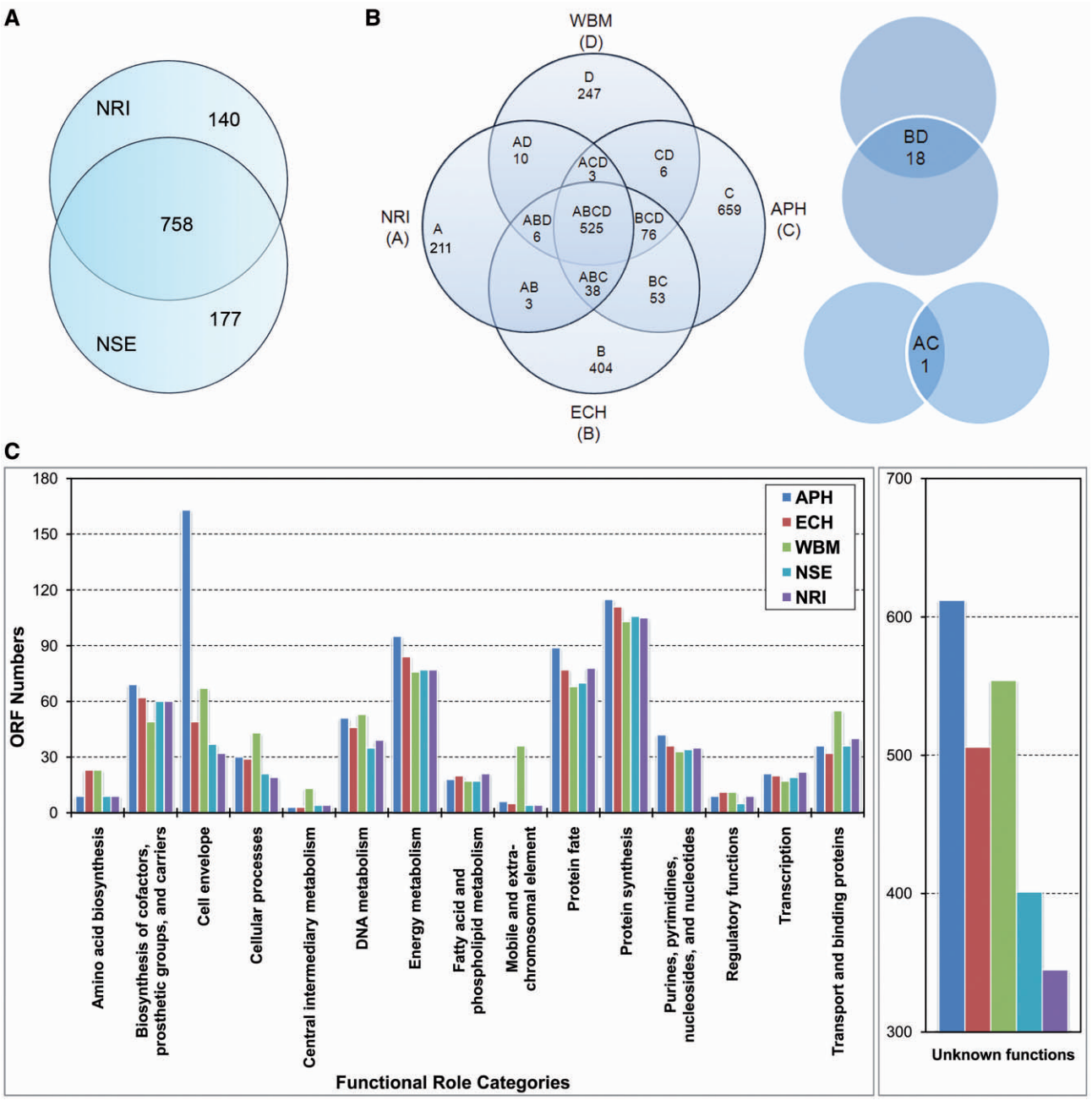

Figure 4. Comparison of the gene sets in members of the family Anaplasmataceae. Venn diagram showing the comparison of conserved and unique genes between *Neorickettsia* spp. (**A**), or among selected members of the family Anaplasmataceae (**B**). Numbers within the intersections of different circles indicate ortholog clusters shared by 2, 3, or 4 organisms. (**C**) Comparison of gene sets by functional role category breakdown. Species indicated in the diagram are as follows: *N. sennetsu* (NSE), *N. risticii* (NRI, A), *E. chaffeensis* (ECH, C), *A. phagocytophilum* (APH, C), and the *Wolbachia* endosymbiont of *Brugia malayi* (WBM, D).

Supplementary Tables 5 and 6). *N. risticii* shares very limited numbers of ortholog clusters (mostly less than 10) in both two- and three-way comparisons (Figure 4B and Supplementary Tables 4–6). The only exception is that, in three-way comparisons, *N. risticii* shares 38 ortholog clusters with *E. chaffeensis* and *A. phagocytophilum*, but not *Wolbachia* wBm (Figure 4B). Most of these orthologs are enzymes involved in vitamin and cofactor (biotin, folate and NAD) biosynthesis (15 orthologs, 40% of 38 orthologs), along with one putative transcriptional regulator (ECH_1118, APH_1218, or NRI_0223), suggesting a potential pathogenic trait or a niche adaption in mammalian hosts (Supplementary Tables 5 and 7). Interestingly, comparison with members of Anaplasmataceae shows *N. risticii* and *N. sennetsu* possess some genes that have either no homology or low homology to those in Anaplasmataceae, but higher homology to those of γ-proteobacteria, such as HU DNA-binding proteins, ATP synthase subunits, DsbB/D protein, and some ribosomal protein subunits.

Table 3. Role category breakdown of conserved or unique *N. risticii* genes by two- and four-way comparisons

Role category	Total genes	Conserved genes (four-way)[a]	Unique genes (two-way w/NSE)[b]
Amino-acid biosynthesis	9	7	
Biosynthesis of cofactor and vitamin	60	40	
Cell envelope	32	10	1[c]
Cellular processes	19	15	
Central intermediary metabolism	4	4	
DNA metabolism	39	35	1[d]
Energy metabolism	77	69	
Fatty acid and phospholipid metabolism	21	19	
Mobile elements	4	1	
Protein fate	78	63	
Protein synthesis	105	98	
Nucleotide biosynthesis	35	33	
Regulatory functions	9	8	
Transcription	22	19	
Transport and binding proteins	40	28	2[e]
Unknown functions	344	76	120
Total numbers	898	525	124

[a]Conserved genes of *N. risticii* are determined based on four-way comparison with other Anaplasmataceae, including *E. chaffeensis*, *A. phagocytophilum*, and the *Wolbachia* endosymbiont of *Brugia malayi*.
[b]*N. risticii*-specific genes are determined based on two-way comparison with *N. sennetsu* (NSE).
[c]*N. risticii* NRI_0742, putative lipoprotein (142 aa, 15.9 kDa).
[d]*N. risticii* NRI_0805, deoxyribodipyrimidine photolyase.
[e]*N. risticii* NRI_0032/NRI_0033, multisubunit Na+/H+ antiporter, MnhB subunits.

As the genomes of *N. risticii*, *N. sennetsu*, *E. chaffeensis*, *A. phagocytophilum* and *Wolbachia* wBm have been annotated and assigned role categories in an identical manner, we compared the number of genes pertaining to different functional roles. This comparison demonstrates that although *N. risticii* and *N. sennetsu* encode fewer proteins, they possess similar numbers of genes involved in housekeeping functions (Figure 4C). The only reduction of gene number occurs in the category of proteins with unknown functions, suggesting that essential housekeeping genes cannot be sacrificed as the genome reduces, including genes encoding metabolism, protein synthesis, transport and regulatory functions (Figure 4C).

Metabolism

Central metabolic pathways and transporters. A previous study showed that *N. risticii* is an aerobic organism and capable of ATP synthesis from glutamine *in vitro* (34). Analysis of the *N. risticii* genome suggests that it encodes pathways for aerobic respiration, including pyruvate metabolism, the tricarboxylic-acid cycle, and the electron transport chain (F_0F_1-ATPase, NADH dehydrogenase, succinate dehydrogenase, cytochrome reductase, and cytochrome oxidase complexes). However, unlike *Rickettsia prowazekii* and *R. conorii* (35), *N. risticii* does not encode d-type oxidase (*cydAB*, cytochrome d ubiquinol oxidase), which is expressed during low oxygen conditions and has a higher affinity for oxygen. *N. risticii* is unable to use glucose, fructose, or fatty acids directly as a carbon or energy source, since essential enzymes for the utilization of these substrates are not identified. The *N. risticii* genome only encodes a partial glycolysis pathway, which lacks hexokinases and is thus unable to convert glucose to glucose-6-phosphate, and a partial gluconeogenesis pathway, which terminates at fructose-6-phosphate. The enzymes in the glycolysis pathway are limited to those that produce glyceraldehydes-3-phosphate and dihydroxyacetone phosphate from phosphoenolpyruvate. A complete nonoxidative pentose-phosphate pathway exists in *N. risticii*, which utilizes glyceraldehydes-3-phosphate to produce pentose that is needed for nucleotide and cofactor biosynthesis. To compensate for the absent or incomplete pathways, the *N. risticii* genome contains several orthologs involved in membrane transport systems that can supply the necessary amino acids, metabolites, and ions (Supplementary Table 8) (36,37). Unlike *Rickettsia* spp., *N. risticii* does not encode translocases for ATP or NADH (ATP:ADP antiporter family), so it appears to rely on its own intracellular ATP production or have a unique ATP acquisition mechanism (Supplementary Table 8).

Nucleotide, cofactor and amino-acid biosynthesis. Like all other Anaplasmataceae (17), *N. risticii* is limited to synthesizing only a few amino acids, such as aspartate, glycine and glutamine and thus must transport most amino acids from its host (Supplementary Table 7). However, *N. risticii* is capable of synthesizing all nucleotides and most vitamins or cofactors, such as biotin, folate, FAD, NAD, CoA and protoheme (Supplementary Table 7). Since *N. risticii* is maintained transstadially and transovarially inside its invertebrate trematode host, it is possible that *N. risticii* is beneficial to the host by providing necessary vitamins and cofactors (mutualism) (38).

Information transfer and DNA repair

Genes-encoding enzymes for DNA replication and repair, RNA synthesis and degradation, ribosomal proteins (except L30 and S22), as well as genes involved in homologous recombination, including RecA/RecF pathways and RuvABC complexes, are all identified in the *N. risticii* genome. In addition, two small RNAs are also found in the genome, including *rnpB* (RNase P RNA component precursor) and tmRNA (responsible for tagging incomplete proteins for proteolysis on stalled ribosomes). A previous study showed that *N. sennetsu* encodes the most limited genes for DNA repair among several sequenced members of the family Anaplasmataceae (17). Similarly, *N. risticii* encodes *uvrD* family genes, but lacks most genes required for the repair of UV-induced DNA damage, including various glycosylases for base excision repair (BER, such as *3mg* and *fpg*) and exonucleases for nucleotide excision repair (NER, such as *uvrABC*). Instead, *N. risticii* might have gained limited DNA repair function by the acquisition of a gene (NRI_0805) most closely related to DNA photolyases of γ-proteobacteria, which is an alternative mechanism to repair UV-damaged

DNA (Supplementary Table 2). The absence of many DNA repair pathways might account for the longer evolutionary branch of *Neorickettsia* spp. in the family Anaplasmataceae (Figure 1).

Despite the distinct trematode hosts for *N. risticii* (which reside in aquatic insects) and *N. sennetsu* (which supposedly reside in fish), there is almost complete genomic synteny (Figure 3). It is assumed that, on average, the number of genome rearrangements increases with time (or sequence divergence) (39). However, synteny analysis between *E. chaffeensis* and *E. canis*, or *A. phagocytophilum* and *A. marginale*, both of which have a small phylogenetic divergence as *N. risticii* and *N. sennetsu* (Figure 1), showed a single symmetrical inversion in the genomes (40). This suggests that there have been very few rearrangement events in these genomes, probably because large rearrangements are lethal for *Neorickettsia* spp. due to minimum sets of mutation repair genes. On the other hand, this poor mutation repair system allows accumulation of point mutations, resulting in new genes with potential novel functions. As housekeeping genes are essential to *Neorickettsia* spp., only those mutations that do not affect housekeeping functions accumulate. A poor mutation repair system due to the loss of mutation repair genes is an important virulence factor of common pathogens such as *Salmonella* and *Escherichia coli* (41). Therefore, the limited DNA repair system of *Neorickettsia* spp. may provide species survival advantage by allowing the accumulation of permissive mutations, thus enhancing genetic diversity in the stable intra-trematode environment where events leading to genomic rearrangements may be rare.

Pathogenesis

Little is known about the genetic determinants required for *N. risticii* to invade hosts and cause disease. Here, we identify several putative genes that might be involved in *N. risticii* pathogenesis, including outer membrane proteins, protein secretion systems, two-component regulatory systems, and other *N. risticii*-specific genes.

Protein secretion systems. Extracellular secretion of products is the major mechanism by which pathogenic Gram-negative bacteria alter host cell functions, thus enhancing survival of the bacteria and/or damaging hosts. In order to secrete virulence factors across two lipid bilayers and the periplasm in between, Gram-negative pathogens use at least six distinct extracellular protein secretion systems, referred to as type I–VI secretion systems or T1SS-T6SS (42). Analysis of the *N. risticii* genome shows that it encodes both Sec-dependent and Sec-independent protein export pathways for secretion of proteins across the membranes, but no homologs of T2SS, T3SS, T5SS or T6SS components could be found in the *N. risticii* genome (43–45). All major components for the Sec-dependent pathway are identified, including the cytosolic protein-export chaperone SecB and the preprotein translocase complex SecY/SecG/SecA/SecD/YajC. The Sec-independent T1SS, an ATP-driven ABC transporter system, is capable of transporting target proteins carrying a C-terminal secretion signal across both inner and outer membranes and into the extracellular medium. Both components of the T1SS, the ATPase AprD and the membrane fusion HlyD family protein AprE, are identified in the *N. risticii* genome.

The twin-arginine dependent translocation (Tat) pathway can transport folded proteins across the bacterial cytoplasmic membrane by recognizing N-terminal signal peptides harboring a distinctive twin-arginine motif (46,47). Components of the Tat pathway including TatA and TatC are also identified in the *N. risticii* genome. Using the TATFIND algorithm (http://signalfind.org/tatfind.html), prokaryotic Tat signal peptides are identified in two *N. risticii* proteins (NRI_0595 and NRI_0661). Nevertheless, the presence of a functional Tat pathway in *N. risticii* remains to be studied. In addition, a set of common chaperonins is also identified in *N. risticii* genome, including *groEL*, *groES*, *dnaK*, *dnaJ*, *hscB* and *htpG*.

The T4SS is a protein secretion system that can translocate bacterial effector molecules into host cells and plays a key role in pathogen-host cell interactions, usually involved in pathogenesis of Gram-negative bacteria (48). In several intracellular bacteria including the family Anaplasmataceae, the T4SS is essential for survival inside host cells, involved in nutrient acquisition, inhibition of phago-lysosomal fusion pathways, and most importantly, establishment of intracellular compartments and intracellular survival (23,49,50). In the *N. risticii* genome, we also identified a T4SS encoded by *virB/D* genes distributed in four separate loci: *virB8-2/9-2*, *virB8-1/virB9-1/virB10/virB11/virD4*, *virB2-1/virB2-2/virB4-2* and *virB3/virB4-1/virB6-1/virB6-2/virB6-3/virB6-4* (Figure 5A). The organization of *virB/D* gene clusters is conserved between *N. risticii* and *N. sennetsu*, but is shifted and/or inverted relative to that of *E. chaffeensis* and *A. phagocytophilum* (Figure 5A). In addition, there are a number of differences. Unlike *A. phagocytophilum* and *E. chaffeensis* (51), the *sodB* gene is not present upstream of the *virB3/virB4/virB6* operon in *Neorickettsia* spp., and *virB8-2/9-2* are organized into a cluster. Furthermore, *N. risticii* encodes a much shorter *virB6-4* (~1000 aa) compared to those of other Anaplasmataceae (>2000 aa). Analysis of *N. risticii* and *N. sennetsu* genome sequences shows that they both encode two copies of *virB2* downstream of *virB4* (Figure 5). Phylogenic analysis shows that *virB2* genes of *N. risticii* and *N. sennetsu* are closely related to *virB2* genes of other α-proteobacteria, but are phylogenetically distinct from *virB2* genes of *E. chaffeensis* and *A. phagocytophilum*, which form a separate clade (Figure 5B) (32).

Subcellular fractionation and functional studies have demonstrated that VirB2 is the major pilus component of T4SS extracellular filaments and may play a critical role in mediating the initial interaction with host cells (48,49). RT–PCR and western blotting results indicate that VirB2 is expressed by *N. risticii* in infected $P388D_1$ cells, and two *virB*2 genes can be co-transcribed as an operon (Figure 6A and B). Double immunofluorescence assays from both intracellular *N. risticii* organisms in permeabilized infected cells (Figure 6C)

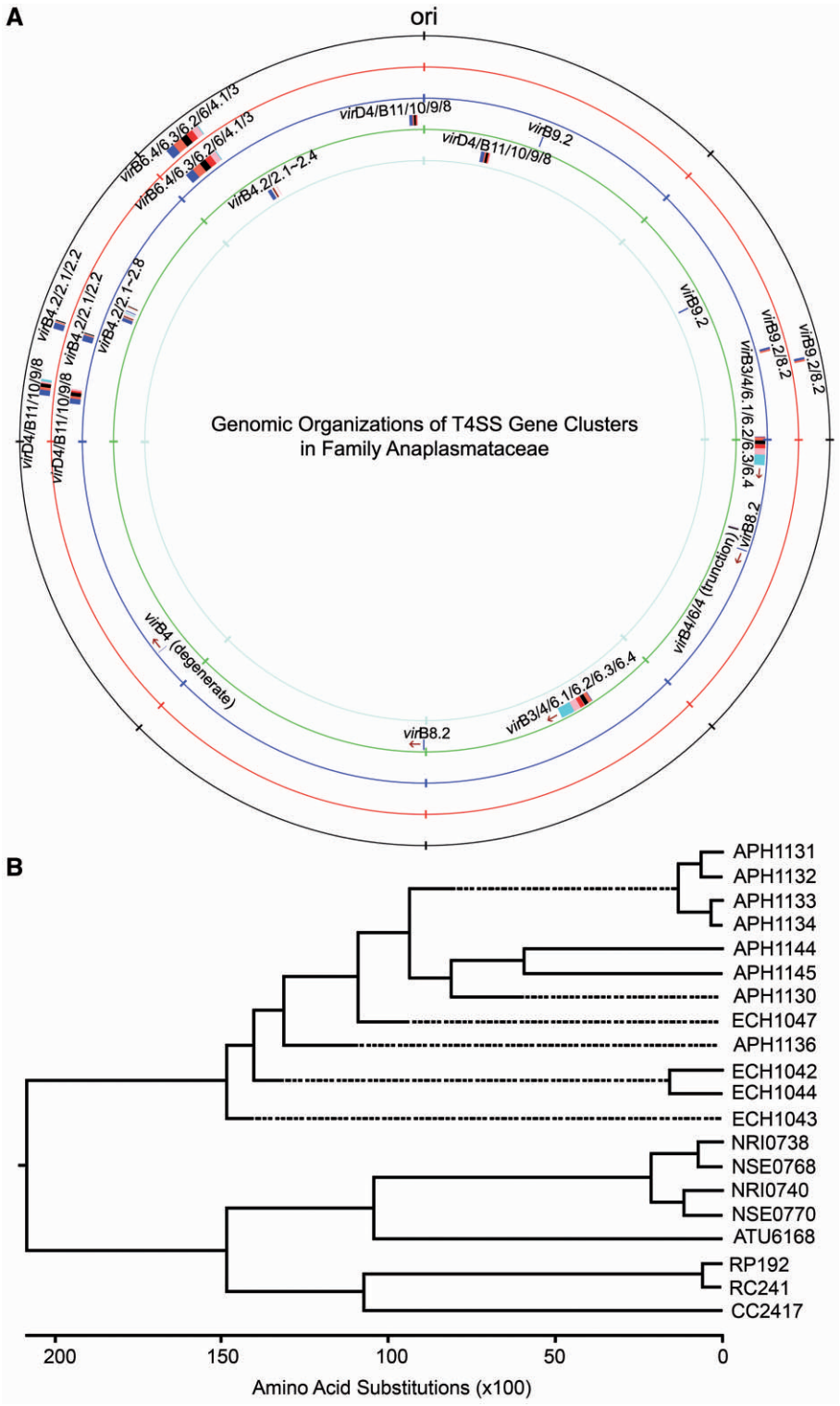

Figure 5. Genomic organization of T4SS *virB*/*D* clusters and phylogenic analysis of *virB2* genes in the family Anaplasmataceae. (**A**) Circular representation of T4SS *virB*/*D* genes in genomes of the family Anaplasmataceae. From outside to inside circles: first (black)—*N. risticii* (879 977 bp); second (red)—*N. sennetsu* (859 006 bp); third (blue)—*A. phagocytophilum* (1 471 282 bp); fourth (green)—*E. chaffeensis* (1 176 248 bp). The individual *virB*/*D* genes were color-coded in the clusters for better visualization. Note that all T4SS genes of *N. risticii* and *N. sennetsu*, and most T4SS genes of *A. phagocytophilum* and *E. chaffeensis* are encoded in the '−' strand, therefore, the location of most genes appear close to the inner circles. Genes encoded in the '+' strand are followed by red arrows. (**B**) Phylogenetic tree of *virB2* genes in the family Anaplasmataceae. Nucleotide sequences of *virB2* from members of the family Anaplasmataceae were aligned using the Clustal W method, and a phylogenetic tree was built. APHxxxx: locus ids for *virB2* genes of *A. phagocytophilum* HZ; ECHxxxx: locus ids of *E. chaffeensis* Arkansas; NSExxxx: locus ids of *N. sennetsu* Miyayama; NRIxxxx: locus ids of *N. risticii* Illinois; ATU6168: *Agrobacterium tumefaciens* C58 pilin subunit, TrbC/VirB2 family protein (Accession No. NP_396488); RP192: *Rickettsia prowazekii* Madrid E TrbC/VirB2 family protein (Accession No. NP_359878); RC241: *Rickettsia conorii* Malish 7 TrbC/VirB2 family protein (Accession No. NP_359878); CC2417: *Caulobacter crescentus* CB15 TrbC/VirB2 family protein (Accession No. NP_421220).

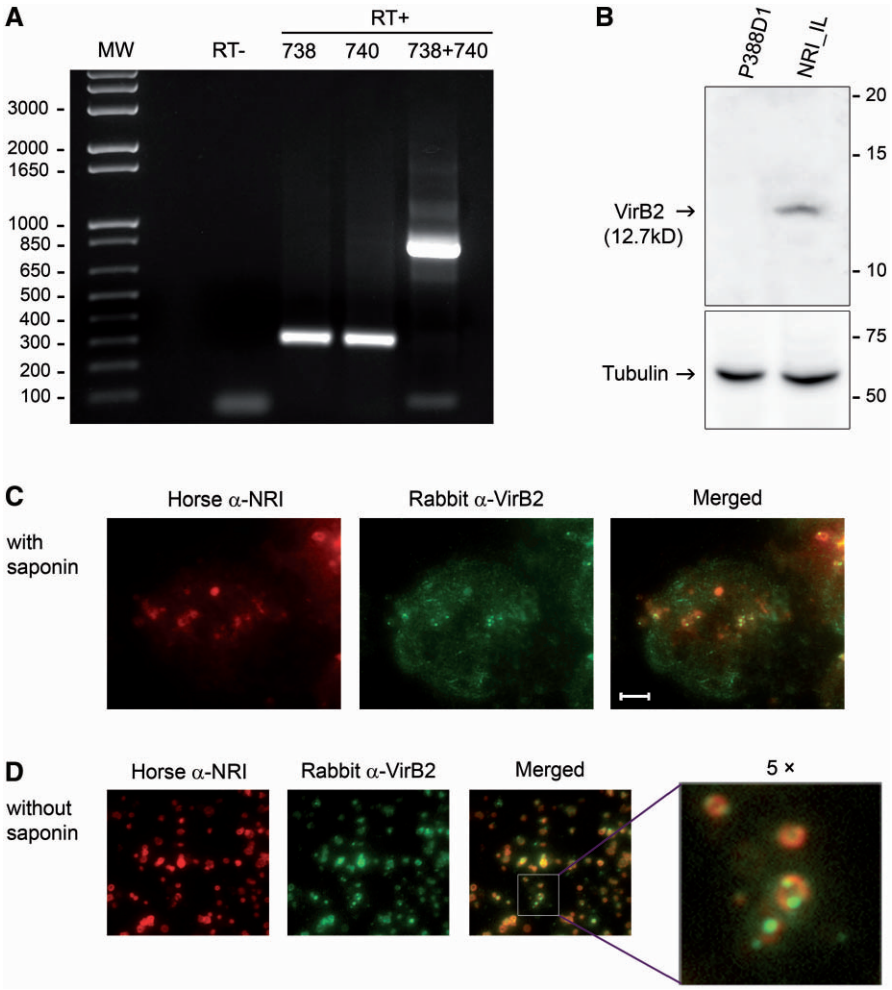

Figure 6. Expression and localization of *N. risticii* VirB2 (NRI_0738/NRI_0740). (**A**) RNAs were prepared from *N. risticii*-infected P388D$_1$ cells at 3 days p.i. DNase-treated total RNA was reverse-transcribed (RT+) and subsequently PCR-amplified using primers specific to *virB2* genes: NRI_0738 and NRI_0740. RT–: negative control without reverse transcriptase; 738: PCR with primers targeting NRI_0738 only; 740: PCR with primers targeting NRI_0740 only; 738 + 740: PCR with primers spanning NRI_0738 and NRI_0740. The relative sizes of molecular mass standards are shown (in base pairs) on the left. (**B**) Whole-cell lysates from *N. risticii*-infected P388D$_1$ cells (3 days p.i.) were prepared and subjected to western blotting using antibody against *N. risticii* VirB2 (NRI_0738), or host α-tubulin as a loading control. The molecular mass standards are shown (in kDa) on the right. (**C**) *N. risticii*-infected P388D$_1$ cells at 3 days p.i. were fixed, permeabilized with 0.1% saponin and dual labeled with horse anti-*N. risticii* (NRI) and rabbit anti-VirB2-1 (NRI_0738) antibodies. Results are representative of three independent experiments. Note bright dots of VirB2 on most of intracellular bacteria. Bar: 5 μm. (**D**) Host cell-free *N. risticii* purified from infected P388D$_1$ cells at 3 days p.i. were fixed and dual labeled with horse anti-*N. risticii* (NRI) and rabbit anti-VirB2 antibodies. The right panel is a 5× amplification of the inlet from the merged image. Results are representative of three independent experiments. Note average bipolar localization of VirB2 per bacterium on the surface of *N. risticii* and variable sizes of bacteria.

and non-permeabilized purified *N. risticii* organisms show that VirB2 forms a punctate pattern, mostly localized at two poles on the surface of *N. risticii* (Figure 6D). These results suggest that VirB2 can form focal complexes on the surface of the *N. risticii* organism, which might serve as channels of the T4SS apparatus, like that of *Agrobacterium* (48).

Two-component and one-component regulatory systems. A two-component regulatory system (TCS) is a signal transduction system that allows bacteria to sense and respond rapidly to changes in their environment. Through the regulation of downstream gene expression, the TCS plays a key role in controlling virulence responses of a wide variety of bacterial pathogens (52). A TCS is generally composed of a sensor histidine kinase and a cognate response regulator, which is often a transcription factor and controls gene expression in response to environmental changes. Our previous studies showed that *E. chaffeensis* and *A. phagocytophilum* express three pairs of TCS, including CckA/CtrA, PleC/PleD and NtrX/NtrY and that their histidine kinase activities were required for bacterial infection (53,54). Computational analysis reveals that the *N. risticii* genome only encodes two pairs of TCS: CckA/CtrA and PleC/PleD (Table 4). The response regulator CtrA is a global transcription factor found only in α-proteobacteria that coordinates multiple cell cycle events at the transcriptional level. The other regulator PleD contains a GGDEF domain, which is associated with diguanyl cyclase activity that generates

Table 4. Potential pathogenic genes and protein locations in *N. risticii* and other Anaplasmataceae

Organisms[a]	NRI	NSE	APH	ECH	WBM
Type IV secretion system					
VirB1	−	−	−	−	−
VirB2	+ (2)	+ (2)	+ (8)	+ (4)	+ (4)
VirB3	+	+	+	+	+
VirB4	+ (2)	+ (2)	+ (2)	+ (2)	+ (2)
VirB5	−	−	−	−	−
VirB6	+ (4)	+ (4)	+ (4)	+ (4)	+ (4)
VirB7	−	−	−	−	−
VirB8	+ (2)	+ (2)	+ (2)	+ (2)	+ (2)
VirB9	+ (2)	+ (2)	+ (2)	+ (2)	+ (2)
VirB10	+[b]	+	+	+	+
VirB11	+	+	+	+	+
VirD4	+	+	+	+	+
Two-component systems					
PleC/PleD[c]	+	+	+	+	+
NtrY/NtrX	−	−	+	+	−
CckA/CtrA	+	+	+	+	+
One-component systems[d]					
DNA-binding protein	+ (2)	+ (2)	+ (2)	+ (2)	−
MerR transcriptional regulator	+	+	+	+	+
Rrf2/aminotransferase	+	+	+	+	−
dGTP triphosphohydrolase	+	+	−	+	+
EAL domain protein	+	+	−	−	−
Ankyrin-repeat domain protein	3	3	4	4	21
Protein locations[e]					
Extracellular	1	1	2	2	1
Outer membrane	8	7	21	22	16
Periplasmic	4	3	8	4	4
Cytoplasmic membrane	149	143	156	158	130
Cytoplasmic	302	311	413	271	290
Unknown	434	467	664	648	364

APH, *A. phagocytophilum*; ECH, *E. chaffeensis*; NSE, *N. sennetsu*; NRI, *N. risticii*; WBM, the *Wolbachia* endosymbiont of *Brugia malayi*.
[a]Numbers inside parentheses indicate the copy numbers of the genes; otherwise, only a single copy is present.
[b]*N. risticii* encodes an additional truncated *virB10* gene resulting from a frameshift.
[c]*N. risticii* and *N. sennetsu* encode two copies of PleC that are homologous to PleC of *E. chaffeensis* or *A. phagocytophilum* (*e*-value <*e*–27 and *e*–66, respectively).
[d]The presence of genes encoding one-component regulatory systems was based on Microbial Signal Transduction Database (http://genomics.ornl.gov/mist).
[e]Location of outer member proteins is predicted by the pSort-B algorithm.

cyclic diguanylate (c-di-GMP) to regulate biofilm or extracellular matrix formation (55).

Prokaryotes also carry out signaling events using the one-component regulatory system, which consists of a single protein containing both input and output domains, but lacks the phospho-transfer domains of TCS (56,57). The number of one-component systems and TCS per genome positively correlates with the genome size and the total number of genes (56). Comparison among members of the family Anaplasmataceae showed that, although *N. risticii* has the smallest genome size, it encodes the highest numbers of one-component system proteins (Table 4 and Supplementary Figure 3). The disproportionate increase in the number of one-component systems might result from the complex lifecycle and environmental conditions of *N. risticii* (56).

Interestingly, analysis of the *N. risticii* genome shows that it encodes two copies of PleC histidine kinase (Table 4) and a one-component signal transduction protein, an EAL domain protein (NRI_0417) (Supplementary Figure 3) (57). An EAL domain protein has diguanylate phosphodiesterase (PDE) activity and can function as a c-di-GMP PDE that converts c-di-GMP to GMP. EAL domain proteins are absent in *A. phagocytophilum* and *E. chaffeensis* (58,59).

Transcriptional regulations. Members in the order Rickettsiales have a small number of transcriptional regulators; a trait also observed in other intracellular pathogens. This phenomenon appears to be the result of reductive evolution coupled with the diminished demand for regulation due to the small genome size, as well as the relatively stable conditions provided by the intracellular environment of the host cells. Like all Anaplasmataceae, *N. risticii* encodes only two sigma factors: RNA polymerase sigma-70 factor (RpoD), which is the essential primary sigma factor and responsible for most RNA synthesis in exponentially growing cells, and sigma-32 factor (rpoH), which is responsible for expression from heat shock promoters. In addition, the response regulator CtrA might regulate the expression of genes involved in the bacterial developmental cycle. Our previous studies demonstrated that *A. phagocytophilum* and *E. chaffeensis* encode a transcriptional regulator *ApxR* and *EcxR*, respectively, which regulates the expression of genes encoding P44 outer membrane proteins and the T4SS (60–62). Based on sequence homology, the ortholog *N. risticii* transcriptional regulator (*NrxR*, NRI_0036), which encodes a 12.5-kDa DNA binding protein, is identified.

Ankyrin domain proteins. Ankyrin repeats have been found in numerous proteins mediating specific protein–protein interactions in bacteria and eukaryotes. Our previous study has suggested that the ankyrin repeat-containing protein AnkA of *A. phagocytophilum* is secreted into host cells by the T4SS and plays an important role in facilitating intracellular infection by activating the Abl-1 protein tyrosine kinase signaling pathway (23). Three proteins that contain ankyrin-repeat domains are identified in the *N. risticii* genome. However, whether any of the ankyrin repeat-containing proteins can be secreted by the T4SS and regulate host-cell signaling remains to be studied.

Cell-wall components

Lipopolysaccharide and peptidoglycan. Most of the genes encoding peptidoglycan and lipopolysaccharide (LPS), including lipid A are absent in members of the family Anaplasmataceae (17,63). The *N. risticii* genome does not encode any enzymes involved in the biosynthesis of lipid A (the core component of LPS) and lacks most enzymes required for the biosynthesis and degradation of diaminopimelate and murein sacculus (components of peptidoglycan) indicating that LPS and peptidoglycan are

absent in *N. risticii*. Since conserved pathogen-associated molecular patterns like LPS or peptidoglycan are potent stimulants for innate immunity and anti-microbial responses in mammalian host immune defensive cells (64,65), the loss of LPS and peptidoglycan eliminates these microbicidal activities in leukocytes and enhances survival for *N. risticii* in mammalian monocytes/macrophages.

Lipoproteins. Studies have suggested that *E. chaffeensis* is capable of expressing mature lipoproteins (66), which involves three lipoprotein-processing enzymes: a prolipoprotein diacylglyceryl transferase (Lgt) that attaches the thiol group of the conserved cysteine residue in the lipobox with a diacylglyceryl group, a lipoprotein-processing protease signal peptidase II (SPase II, LspA) that cleaves prolipoproteins before the conserved cysteine residue in the lipobox, and an apolipoprotein *N*-acyltransferase (Lnt) that adds lipids to the amine of the newly cleaved amino-terminal cysteine (cylation) (67,68). All three lipoprotein-processing enzymes are identified in *N. risticii*: *lgt* (NRI_0823), *lspA* (NRI_0879), and *lnt* (NRI_0470). Furthermore, thirteen lipoproteins are identified by computational prediction using the LipoP 1.0 program (http://www.cbs.dtu.dk/services/LipoP), which is specific for the prediction of lipoproteins and signal peptides in Gram-negative bacteria (Supplementary Table 9) (69). Among these lipoproteins, nine are predicted to be located on the outer membrane. Given the role of *E. chaffeensis* lipoproteins in inducing delayed-type hypersensitivity reaction in dogs (66), it is likely that *N. risticii* lipoproteins are also involved in pathogenesis and immune response in infected horses.

Putative outer membrane proteins. Computational analysis with the pSort-B algorithm predicted nine outer membrane proteins, two of which (rotamase family protein and outer membrane protein assembly complex YaeT protein or Omp85 family protein) might be involved in outer membrane protein transport and stabilization (70) (Table 4 and Supplementary Table 10). Unlike most other members in the family Anaplasmataceae, which encode a diverse complement of OMP-1/MSP2/P44 outer member superfamily proteins, *N. risticii* encodes only one group of putative outer surface proteins that falls into this PFAM family (Pfam01617) (17). This group of proteins consists of three *Wolbachia* surface protein (WSP)-like *N. risticii* surface proteins, each approximately 30-kDa in mass (NSPs: NRI_0838, NRI_0839, and NRI_0841) (Figure 7A–C, and Supplementary Table 10).

WSPs have been identified as highly expressed proteins in *Wolbachia* strains and are likely surface-exposed, based on their similarities to other outer surface proteins in related bacteria (such as MSP4 in *A. marginale*) and through membrane fractionation and bacterial labeling experiments (71,72). Unlike WSP paralogs of *Wolbachia* strains that are scattered in the genome (38), *N. risticii nsp* genes (NRI_0838, NRI_0839 and NRI_0841) comprise a potential operon (Figure 7A). Alignment of *Neorickettsia nsp* and *Wolbachia wsp* genes shows that they are well conserved, with more than 35% identity among

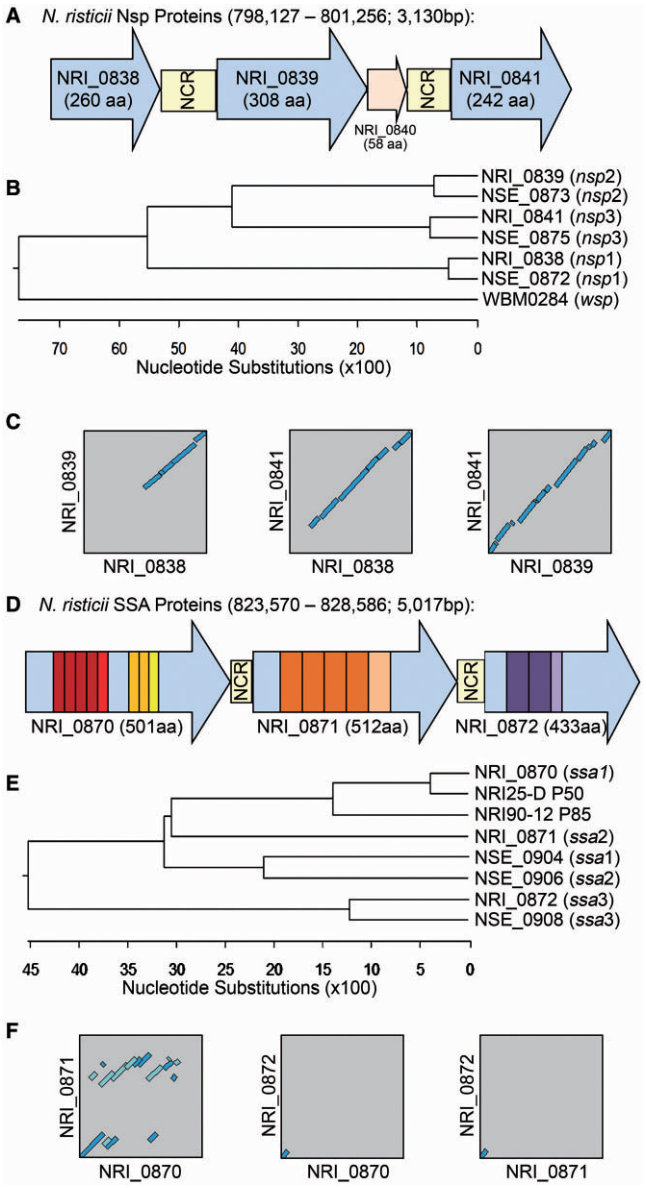

Figure 7. Genomic organization and alignment of *N. risticii* putative membrane proteins NSP and SSA. (**A** and **D**) Genomic organization of *N. risticii* genes encoding putative membrane proteins NSP (A) and SSA (D). Color bars inside SSA genes indicate repeat sequences. NCR: non-coding region. (**B** and **E**) Phylogenetic tree of *nsp* (B) and *ssa* (E) genes in *N. risticii* and *N. sennetsu*. Nucleotide sequences of *nsp* or *ssa* from *N. risticii* and *N. sennetsu* were aligned using the Clustal W method, and a phylogenetic tree was built. NRI25-D: *N. risticii* strain 25-D; NRI90-12: *N. risticii* strain 90-12. (**C** and **F**) Alignment of *N. risticii* NSP (C) and SSA (F) protein sequences using blast2seq.

*Wolbachia wsp*s, *N. risticii nsp*s and *N. sennetsu nsp*s (which were previously annotated to encode hypothetical proteins) (17) (Figure 7B). Furthermore, alignment of three *N. risticii* NSP proteins using blast2seq shows that NSP2 (NRI_0839) is related to NSP3 (NRI_0841), but NSP2 or NSP3 only share C-terminal fragments with NSP1 (NRI_0838) (Figure 7C). *Wolbachia* WSP proteins are known to inhibit human neutrophil apoptosis, and induce inflammatory responses in human neutrophils (73–75). It is possible that *N. risticii* NSP proteins can

act in a similar fashion to *Wolbachia* WSP proteins, regulating host cell functions and immune responses. In addition, two members in this family, *A. phagocytophilum* P44 and *E. chaffeensis* P28/OMP-1F, have been demonstrated to possess porin activities of relatively large pore size to allow passive diffusion of L-glutamine, monosaccharides, disaccharides, and even tetrasaccharides (76,77). Given the similar predicted structures of amphipathic and antiparallel β-strands, abundant polar residues, and a phenylalanine residue at the C-terminus of both NSP1 and NSP2, *N. risticii* NSP might also have porin activity to facilitate bacterial intracellular growth.

Computational analysis of the *N. risticii* genome and recent molecular characterization of *N. risticii* have identified alternative sets of potential surface proteins, which include one 51-kDa protein (P51), and three strain-specific antigens (SSA). P51 is a specific antigenic protein conserved among *N. sennetsu*, *N. risticii* and *Stellantchasmus falcatus* trematode agent (SF agent) (78). A previous study had shown that P51 is the major antigen recognized in horses with PHF (79). Computational analysis by SignalP indicates that P51 encodes a signal peptide, and an immunofluorescence assay using anti-P51 antibody on fixed but non-permeabilized *N. risticii* organisms showed a ring-like labeling pattern surrounding the bacteria, supporting bacterial surface exposure of P51 (80). Jaccard-filtered COG analysis indicates that P51 belongs to an ortholog cluster that presents in all Rickettsiales (*E*-value $<1e{-}5$) (17), and computational prediction using PRED-TMBB method (http://biophysics.biol.uoa.gr/PRED-TMBB/) shows 18 transmembrane β strands in the P51 protein (81). The discrimination score of P51 for β-barrel proteins is 2.910, lower than the threshold value of 2.965, suggesting that it is a β-barrel outer membrane protein (Supplementary Figure 4).

Strain-specific antigens (SSAs), which encode proteins at 50- or 85-kDa in different strains of *N. risticii*, were initially identified during cross-reaction studies between strains 25-D (P50/SSA) and 90-12 (P85/SSA) of *N. risticii* isolated from horses with PHF (15,82). Surface labeling with ^{125}I suggested that P50/SSA from *N. risticii* strain 25-D is a surface antigen (83). Furthermore, P50/SSA has been determined to be a protective antigen of *N. risticii* against homologous challenge; however, the length and number of repeats varied within SSAs of each strain, which is suggested to contribute to vaccine failure (15,82). Analysis of the *N. risticii* Illinois genome identifies three SSA proteins (NRI_0870, NRI_0871 and NRI_0872), which apparently are arranged in a single operon (Figure 7D). Sequence analysis indicates that the three *N. risticii* SSAs contain different numbers of intramolecular tandem repeats, which may contribute to the antigenic variation of this pathogen (Figure 7D, colored boxes in the ORF, and Supplementary Table 11) (84). Alignment of *Neorickettsia* SSA genes or proteins using Clustal W or blast2seq shows that *ssa1* (NRI_0870) is closely related to *ssa2* (NRI_0871); however, *ssa3* (NRI_0872) only shares a short N-terminal fragment with either *ssa1* or *ssa2* and belongs to a separate clade from *ssa1* and *ssa2* (Figure 7E and F). Proteins with tandem repeats play important roles in pathogenesis and pathogen–host interactions, including creating phase variations (Vlp of *Mycoplasma hyorhinis*) and controlling bacterial motilities (ActA of *Listeria monocytogenes*) (85,86). In additional to SSA proteins, intragenic tandem repeats are identified in 17 proteins of *N. risticii* (Supplementary Table 12). Among them, 11 proteins are predicted to localize on bacterial membrane or periplasm.

The roles of these putative outer membrane proteins in infecting horses and regulating host immune responses remain to be elucidated; nevertheless, these proteins could be interesting candidates for protective PHF vaccine development.

CONCLUSIONS

Analysis of the genomic sequence of *N. risticii* provides the resources necessary for a detailed study of this bacterium within trematodes hosts and for comprehensive understanding of molecular pathogenesis in mammalian hosts. Despite apparent divergent trematode hosts: one encysting within insects in North America and another encysting within fish in Japan and Southeast Asia, *N. risticii* and *N. sennetsu* have preserved complete genome synteny, but genome contents are divergent on proteins with unknown functions. These unique hypothetical proteins could be factors that define their host/vector specificities and could be involved in pathogenesis. Surface-exposed proteins and unique sets of proteins involved in pathogenesis identified in the present study provide valuable insights into future research on PHF vaccines and development of novel therapeutic modalities.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACCESSION NUMBER

CP001431.

ACKNOWLEDGEMENTS

The authors would like to thank Dr K. Miura for the initial assembly of the *N. risticii* genome sequence, J. Craig Venter Institute (JCVI) for providing the JCVI Annotation Service including automatic annotation data and the manual annotation tool Manatee, and Dr I.T. Paulsen for transporter analysis.

FUNDING

Ohio State University College of Veterinary Medicine Equine Research grant; partly by R01 AI47885 from the National Institutes of Health. Funding for open access charge: R01 AI 47885.

Conflict of interest statement. None declared.

REFERENCES

1. Rikihisa,Y. and Perry,B.D. (1985) Causative ehrlichial organisms in Potomac horse fever. *Infect. Immun.*, **49**, 513–517.
2. Rikihisa,Y., Perry,B.D. and Cordes,D.O. (1985) Ultrastructural study of ehrlichial organisms in the large colons of ponies infected with Potomac horse fever. *Infect. Immun.*, **49**, 505–512.
3. Knowles,R.C., Anderson,C.W., Shipley,W.D., Whitlock,R.H., Perry,B.D. and Davidson,J.D. (1984) *Proceedings of the 29th Annual Convention of the American Association of Enquine Practitioners*. American Association of Equine Practitioners, Lexington, KY, pp. 353–357.
4. Rikihisa,Y., Perry,B. and Cordes,D. (1984) Rickettsial link with acute equine diarrhoea. *Vet. Rec.*, **115**, 390.
5. Holland,C.J., Ristic,M., Cole,A.I., Johnson,P., Baker,G. and Goetz,T. (1985) Isolation, experimental transmission, and characterization of causative agent of Potomac horse fever. *Science*, **227**, 522–524.
6. Rikihisa,Y., Dumler,J.S. and Dasch,G.A. (2005) Neorickettsia. In Brenner,K., Staley and Garrity (eds), *The Proteobacteria, Part C, Bergey's Manual of Systematic Bacteriology*, Vol. 2, 2nd edn. Springer, New York, NY, pp. 132–137.
7. Dutra,F., Schuch,L.F., Delucchi,E., Curcio,B.R., Coimbra,H., Raffi,M.B., Dellagostin,O. and Riet-Correa,F. (2001) Equine monocytic Ehrlichiosis (Potomac horse fever) in horses in Uruguay and southern Brazil. *J. Vet. Diagn. Invest.*, **13**, 433–437.
8. Gibson,K.E., Rikihisa,Y., Zhang,C. and Martin,C. (2005) Neorickettsia risticii is vertically transmitted in the trematode *Acanthatrium oregonense* and horizontally transmitted to bats. *Environ. Microbiol.*, **7**, 203–212.
9. Gibson,K.E. and Rikihisa,Y. (2008) Molecular link of different stages of the trematode host of *Neorickettsia risticii* to *Acanthatrium oregonense*. *Environ. Microbiol.*, **10**, 2064–2073.
10. Rikihisa,Y. (2006) New findings on members of the family Anaplasmataceae of veterinary importance. *Ann. NY Acad. Sci.*, **1078**, 438–445.
11. Rikihisa,Y., Dumler,J.S. and Dasch,G.A. (2004) Rickettsial diseases. In Reed,S.M., Bayly,W. and Sellon,D.C. (eds), *Equine Internal Medicine*, 2nd edn. W.B. Saunders, Philadelphia, PA, pp. 96–109.
12. Mott,J., Rikihisa,Y., Zhang,Y., Reed,S.M. and Yu,C.Y. (1997) Comparison of PCR and culture to the indirect fluorescent-antibody test for diagnosis of Potomac horse fever. *J. Clin. Microbiol.*, **35**, 2215–2219.
13. Madigan,J.E., Rikihisa,Y., Palmer,J.E., DeRock,E. and Mott,J. (1995) Evidence for a high rate of false-positive results with the indirect fluorescent antibody test for *Ehrlichia risticii* antibody in horses. *J. Am. Vet. Med. Assoc.*, **207**, 1448–1453.
14. Chaichanasiriwithaya,W., Rikihisa,Y., Yamamoto,S., Reed,S., Crawford,T.B., Perryman,L.E. and Palmer,G.H. (1994) Antigenic, morphologic, and molecular characterization of new *Ehrlichia risticii* isolates. *J. Clin. Microbiol.*, **32**, 3026–3033.
15. Dutta,S.K., Vemulapalli,R. and Biswas,B. (1998) Association of deficiency in antibody response to vaccine and heterogeneity of Ehrlichia risticii strains with Potomac horse fever vaccine failure in horses. *J. Clin. Microbiol.*, **36**, 506–512.
16. Pretzman,C., Ralph,D., Stothard,D.R., Fuerst,P.A. and Rikihisa,Y. (1995) 16S rRNA gene sequence of *Neorickettsia helminthoeca* and its phylogenetic alignment with members of the genus *Ehrlichia*. *Int. J. Syst. Bacteriol.*, **45**, 207–211.
17. Dunning Hotopp,J.C., Lin,M., Madupu,R., Crabtree,J., Angiuoli,S.V., Eisen,J., Seshadri,R., Ren,Q., Wu,M., Utterback,T.R. *et al.* (2006) Comparative genomics of emerging human ehrlichiosis agents. *PLoS Genet.*, **2**, e21.
18. Rikihisa,Y. (1991) Protection against murine Potomac horse fever by an inactivated *Ehrlichia risticii* vaccine. *Vet. Microbiol.*, **27**, 339–350.
19. Lobry,J.R. (1996) A simple vectorial representation of DNA sequences for the detection of replication origins in bacteria. *Biochimie.*, **78**, 323–326.
20. Lee,Y.S., Kim,T.H., Kang,T.W., Chung,W.H. and Shin,G.S. (2008) WSPMaker: a web tool for calculating selection pressure in proteins and domains using window-sliding. *BMC Bioinformatics*, **9(Suppl. 12)**, S13.
21. Bendtsen,J.D., Nielsen,H., von Heijne,G. and Brunak,S. (2004) Improved prediction of signal peptides: SignalP 3.0. *J. Mol. Biol.*, **340**, 783–795.
22. Krogh,A., Larsson,B., von Heijne,G. and Sonnhammer,E.L. (2001) Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.*, **305**, 567–580.
23. Lin,M., den Dulk-Ras,A., Hooykaas,P.J. and Rikihisa,Y. (2007) *Anaplasma phagocytophilum* AnkA secreted by type IV secretion system is tyrosine phosphorylated by Abl-1 to facilitate infection. *Cell Microbiol.*, **9**, 2644–2657.
24. Ge,Y. and Rikihisa,Y. (2007) Identification of novel surface proteins of *Anaplasma phagocytophilum* by affinity purification and proteomics. *J. Bacteriol.*, **189**, 7819–7828.
25. Massung,R.F., Hiratzka,S.L., Brayton,K.A., Palmer,G.H. and Lee,K.N. (2008) Succinate dehydrogenase gene arrangement and expression in *Anaplasma phagocytophilum*. *Gene*, **414**, 41–48.
26. Fukuda,Y., Nakayama,Y. and Tomita,M. (2003) On dynamics of overlapping genes in bacterial genomes. *Gene*, **323**, 181–187.
27. Blanc,G., Ogata,H., Robert,C., Audic,S., Suhre,K., Vestris,G., Claverie,J.M. and Raoult,D. (2007) Reductive genome evolution from the mother of *Rickettsia*. *PLoS Genet.*, **3**, e14.
28. Kozak,M. (2005) Regulation of translation via mRNA structure in prokaryotes and eukaryotes. *Gene*, **361**, 13–37.
29. Nei,M. and Gojobori,T. (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. *Mol. Biol. Evol.*, **3**, 418–426.
30. Skovgaard,M., Jensen,L.J., Brunak,S., Ussery,D. and Krogh,A. (2001) On the total number of genes and their length distribution in complete microbial genomes. *Trends Genet.*, **17**, 425–428.
31. Ochman,H. (2002) Distinguishing the ORFs from the ELFs: short bacterial genes and the annotation of genomes. *Trends Genet.*, **18**, 335–337.
32. Nelson,C.M., Herron,M.J., Felsheim,R.F., Schloeder,B.R., Grindle,S.M., Chavez,A.O., Kurtti,T.J. and Munderloh,U.G. (2008) Whole genome transcription profiling of *Anaplasma phagocytophilum* in human and tick host cells by tiling array analysis. *BMC Genomics*, **9**, 364.
33. Lin,M., Kikuchi,T., Mottaz-Brewer,H.M., Norbeck,A.D., Paša-Tolić,L. and Rikihisa,Y. (2009) *ASM 109th General Meeting*. Vol. 109, American Society for Microbiology, Philadelphia, PA, p. D-095.
34. Weiss,E., Williams,J.C., Dasch,G.A. and Kang,Y.H. (1989) Energy metabolism of monocytic *Ehrlichia*. *Proc. Natl Acad. Sci. USA*, **86**, 1674–1678.
35. McLeod,M.P., Qin,X., Karpathy,S.E., Gioia,J., Highlander,S.K., Fox,G.E., McNeill,T.Z., Jiang,H., Muzny,D., Jacob,L.S. *et al.* (2004) Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J. Bacteriol.*, **186**, 5842–5855.
36. Ren,Q., Chen,K. and Paulsen,I.T. (2007) TransportDB: a comprehensive database resource for cytoplasmic membrane transport systems and outer membrane channels. *Nucleic Acids Res.*, **35**, D274–D279.
37. Ren,Q. and Paulsen,I.T. (2007) Large-scale comparative genomic analyses of cytoplasmic membrane transport systems in prokaryotes. *J. Mol. Microbiol. Biotechnol.*, **12**, 165–179.
38. Wu,M., Sun,L.V., Vamathevan,J., Riegler,M., Deboy,R., Brownlie,J.C., McGraw,E.A., Martin,W., Esser,C., Ahmadinejad,N. *et al.* (2004) Phylogenomics of the reproductive parasite *Wolbachia pipientis* wMel: a streamlined genome overrun by mobile genetic elements. *PLoS Biol.*, **2**, E69.
39. Belda,E., Moya,A. and Silva,F.J. (2005) Genome rearrangement distances and gene order phylogeny in gamma-Proteobacteria. *Mol. Biol. Evol.*, **22**, 1456–1467.
40. Brayton,K.A., Dark,M.J. and Palmer,G.H. (2009) Anaplasma. In Nene,V. and Kole,C. (eds), *Genome Mapping and Genomics in Animal-Associated Microbes*. Springer, Berlin/Heidelberg, Germany, pp. 85–116.
41. LeClerc,J.E., Li,B., Payne,W.L. and Cebula,T.A. (1996) High mutation frequencies among *Escherichia coli* and *Salmonella* pathogens. *Science*, **274**, 1208–1211.
42. Papanikou,E., Karamanou,S. and Economou,A. (2007) Bacterial protein secretion through the translocase nanomachine. *Nat. Rev. Microbiol.*, **5**, 839–851.

43. Henderson,I.R., Navarro-Garcia,F., Desvaux,M., Fernandez,R.C. and Ala'Aldeen,D. (2004) Type V protein secretion pathway: the autotransporter story. *Microbiol. Mol. Biol. Rev.*, **68**, 692–744.
44. Cianciotto,N.P. (2005) Type II secretion: a protein secretion system for all seasons. *Trends Microbiol.*, **13**, 581–588.
45. Bingle,L.E., Bailey,C.M. and Pallen,M.J. (2008) Type VI secretion: a beginner's guide. *Curr. Opin. Microbiol.*, **11**, 3–8.
46. Sargent,F., Berks,B.C. and Palmer,T. (2006) Pathfinders and trailblazers: a prokaryotic targeting system for transport of folded proteins. *FEMS Microbiol. Lett.*, **254**, 198–207.
47. Lee,P.A., Tullman-Ercek,D. and Georgiou,G. (2006) The bacterial twin-arginine translocation pathway. *Annu. Rev. Microbiol.*, **60**, 373–395.
48. Cascales,E. and Christie,P.J. (2003) The versatile bacterial type IV secretion systems. *Nat. Rev. Microbiol.*, **1**, 137–149.
49. Backert,S. and Meyer,T.F. (2006) Type IV secretion systems and their effectors in bacterial pathogenesis. *Curr. Opin. Microbiol.*, **9**, 207–217.
50. Niu,H., Rikihisa,Y., Yamaguchi,M. and Ohashi,N. (2006) Differential expression of VirB9 and VirB6 during the life cycle of *Anaplasma phagocytophilum* in human leucocytes is associated with differential binding and avoidance of lysosome pathway. *Cell Microbiol.*, **8**, 523–534.
51. Ohashi,N., Zhi,N., Lin,Q. and Rikihisa,Y. (2002) Characterization and transcriptional analysis of gene clusters for a type IV secretion machinery in human granulocytic and monocytic ehrlichiosis agents. *Infect. Immun.*, **70**, 2128–2138.
52. Dorman,C.J., McKenna,S. and Beloin,C. (2001) Regulation of virulence gene expression in *Shigella flexneri*, a facultative intracellular pathogen. *Int. J. Med. Microbiol.*, **291**, 89–96.
53. Cheng,Z., Kumagai,Y., Lin,M., Zhang,C. and Rikihisa,Y. (2006) Intra-leukocyte expression of two-component systems in *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* and effects of the histidine kinase inhibitor closantel. *Cell Microbiol.*, **8**, 1241–1252.
54. Kumagai,Y., Cheng,Z., Lin,M. and Rikihisa,Y. (2006) Biochemical activities of three pairs of *Ehrlichia chaffeensis* two-component regulatory system proteins involved in inhibition of lysosomal fusion. *Infect. Immun.*, **74**, 5014–5022.
55. Tischler,A.D. and Camilli,A. (2004) Cyclic diguanylate (c-di-GMP) regulates *Vibrio cholerae* biofilm formation. *Mol. Microbiol.*, **53**, 857–869.
56. Ulrich,L.E., Koonin,E.V. and Zhulin,I.B. (2005) One-component systems dominate signal transduction in prokaryotes. *Trends Microbiol.*, **13**, 52–56.
57. Ulrich,L.E. and Zhulin,I.B. (2007) MiST: a microbial signal transduction database. *Nucleic Acids Res.*, **35**, D386–D390.
58. Lai,T.H., Kumagai,Y., Hyodo,M., Hayakawa,Y. and Rikihisa,Y. (2009) The *Anaplasma phagocytophilum* PleC histidine kinase and PleD diguanylate cyclase two-component system and role of cyclic Di-GMP in host cell infection. *J. Bacteriol.*, **191**, 693–700.
59. Romling,U. (2009) Cyclic Di-GMP (c-Di-GMP) goes into host cells—c-Di-GMP signaling in the obligate intracellular pathogen *Anaplasma phagocytophilum*. *J. Bacteriol.*, **191**, 683–686.
60. Cheng,Z., Wang,X. and Rikihisa,Y. (2008) Regulation of type IV secretion apparatus genes during *Ehrlichia chaffeensis* intracellular development by a previously unidentified protein. *J. Bacteriol.*, **190**, 2096–2105.
61. Wang,X., Kikuchi,T. and Rikihisa,Y. (2007) Proteomic identification of a novel *Anaplasma phagocytophilum* DNA binding protein that regulates a putative transcription factor. *J. Bacteriol.*, **189**, 4880–4886.
62. Wang,X., Cheng,Z., Zhang,C., Kikuchi,T. and Rikihisa,Y. (2007) *Anaplasma phagocytophilum* p44 mRNA expression is differentially regulated in mammalian and tick host cells: involvement of the DNA binding protein ApxR. *J. Bacteriol.*, **189**, 8651–8659.
63. Lin,M. and Rikihisa,Y. (2003) *Ehrlichia chaffeensis* and *Anaplasma phagocytophilum* lack genes for lipid A biosynthesis and incorporate cholesterol for their survival. *Infect. Immun.*, **71**, 5324–5331.
64. Beutler,B. (2000) Tlr4: central component of the sole mammalian LPS sensor. *Curr. Opin. Immunol.*, **12**, 20–26.
65. Ulevitch,R.J. and Tobias,P.S. (1994) Recognition of endotoxin by cells leading to transmembrane signaling. *Curr. Opin. Immunol.*, **6**, 125–130.
66. Huang,H., Lin,M., Wang,X., Kikuchi,T., Mottaz,H., Norbeck,A. and Rikihisa,Y. (2008) Proteomic analysis of and immune responses to *Ehrlichia chaffeensis* lipoproteins. *Infect. Immun.*, **76**, 3405–3414.
67. Paetzel,M., Karla,A., Strynadka,N.C. and Dalbey,R.E. (2002) Signal peptidases. *Chem. Rev.*, **102**, 4549–4580.
68. Gupta,S.D. and Wu,H.C. (1991) Identification and subcellular localization of apolipoprotein N-acyltransferase in *Escherichia coli*. *FEMS Microbiol. Lett.*, **62**, 37–41.
69. Juncker,A.S., Willenbrock,H., Von Heijne,G., Brunak,S., Nielsen,H. and Krogh,A. (2003) Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.*, **12**, 1652–1662.
70. Surana,N.K., Grass,S., Hardy,G.G., Li,H., Thanassi,D.G. and Geme,J.W. 3rd (2004) Evidence for conservation of architecture and physical properties of Omp85-like proteins throughout evolution. *Proc. Natl Acad. Sci. USA*, **101**, 14497–14502.
71. Braig,H.R., Zhou,W., Dobson,S.L. and O'Neill,S.L. (1998) Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J. Bacteriol.*, **180**, 2373–2378.
72. Brattig,N.W., Bazzocchi,C., Kirschning,C.J., Reiling,N., Buttner,D.W., Ceciliani,F., Geisinger,F., Hochrein,H., Ernst,M., Wagner,H. *et al.* (2004) The major surface protein of *Wolbachia* endosymbionts in filarial nematodes elicits immune responses through TLR2 and TLR4. *J. Immunol.*, **173**, 437–445.
73. Bazzocchi,C., Comazzi,S., Santoni,R., Bandi,C., Genchi,C. and Mortarino,M. (2007) *Wolbachia* surface protein (WSP) inhibits apoptosis in human neutrophils. *Parasite Immunol.*, **29**, 73–79.
74. Sanogo,Y.O., Dobson,S.L., Bordenstein,S.R. and Novak,R.J. (2007) Disruption of the *Wolbachia* surface protein gene wspB by a transposable element in mosquitoes of the *Culex pipiens* complex (Diptera, Culicidae). *Insect Mol. Biol.*, **16**, 143–154.
75. Porksakorn,C., Nuchprayoon,S., Park,K. and Scott,A.L. (2007) Proinflammatory cytokine gene expression by murine macrophages in response to *Brugia malayi Wolbachia* surface protein. *Mediators Inflamm.*, **2007**, 84318.
76. Huang,H., Wang,X., Kikuchi,T., Kumagai,Y. and Rikihisa,Y. (2007) Porin activity of *Anaplasma phagocytophilum* outer membrane fraction and purified P44. *J. Bacteriol.*, **189**, 1998–2006.
77. Kumagai,Y., Huang,H. and Rikihisa,Y. (2008) Expression and porin activity of P28 and OMP-1F during intracellular *Ehrlichia chaffeensis* development. *J. Bacteriol.*, **190**, 3597–3605.
78. Rikihisa,Y., Zhang,C., Kanter,M., Cheng,Z., Ohashi,N. and Fukuda,T. (2004) Analysis of p51, groESL, and the major antigen P51 in various species of *Neorickettsia*, an obligatory intracellular bacterium that infects trematodes and mammals. *J. Clin. Microbiol.*, **42**, 3823–3826.
79. Vemulapalli,R., Biswas,B. and Dutta,S.K. (1998) Cloning and molecular analysis of genes encoding two immunodominant antigens of *Ehrlichia risticii*. *Microb Pathog*, **24**, 361–372.
80. Gibson,K.E. and Rikihisa,Y. (2008) Identification of surface proteins on Neorickettsia risticii. *Proceedings of the ASM 108th General Meeting*. Vol. 108, American Society for Microbiology, Washington, DC, p. 208.
81. Bagos,P.G., Liakopoulos,T.D., Spyropoulos,I.C. and Hamodrakas,S.J. (2004) PRED-TMBB: a web server for predicting the topology of beta-barrel outer membrane proteins. *Nucleic Acids Res.*, **32**, W400–W404.
82. Biswas,B., Vemulapalli,R. and Dutta,S.K. (1998) Molecular basis for antigenic variation of a protective strain-specific antigen of *Ehrlichia risticii*. *Infect. Immun.*, **66**, 3682–3688.
83. Dutta,S.K., Shankarappa,B., Thaker,S.R. and Mattingly-Napier,B.L. (1990) DNA restriction endonuclease cleavage pattern and protein antigen profile of *Ehrlichia risticii*. *Vet. Microbiol.*, **25**, 29–38.
84. Benson,G. (1999) Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.*, **27**, 573–580.
85. Citti,C. and Wise,K.S. (1995) *Mycoplasma hyorhinis* vlp gene transcription: critical role in phase variation and expression of surface lipoproteins. *Mol. Microbiol.*, **18**, 649–660.
86. Smith,G.A., Theriot,J.A. and Portnoy,D.A. (1996) The tandem repeat domain in the *Listeria monocytogenes* ActA protein controls the rate of actin-based motility, the percentage of moving bacteria, and the localization of vasodilator-stimulated phosphoprotein and profilin. *J. Cell Biol.*, **135**, 647–660.