

**New Perspectives on Rickettsial Evolution from New Genome Sequences of *Rickettsia*,  
particularly *R. canadensis*, and *Orientia tsutsugamushi***

**Marina E. Ereemeeva<sup>1#\*</sup>, Anup Madan<sup>2</sup>, Chris Shaw<sup>3</sup>, Kevin Tang,<sup>4#</sup> and Gregory A.  
Dasch<sup>1#</sup>**

<sup>1</sup>Viral and Rickettsial Zoonoses Branch, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333, <sup>2</sup>Laboratory of Neurobiology, University of Iowa, Iowa City, Iowa, <sup>3</sup>School of Interactive Arts and Technology, Simon Fraser University Surrey, Surrey, British Columbia, Canada, and <sup>4</sup>Scientific Resources Program, National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia 30333

\*Corresponding author:

Viral and Rickettsial Zoonoses Branch,  
Mail Stop G-13  
National Center for Infectious Diseases  
Centers for Disease Control and Prevention  
1600 Clifton Road NE  
Atlanta GA 30333  
Phone: 404-639-4612  
Fax: 404-639-4436  
E-mail: [MEreemeeva@cdc.gov](mailto:MEreemeeva@cdc.gov)

<sup>#</sup>The findings and conclusions in this report are those of the authors and do not necessarily represent the views of the Centers for Disease Control & Prevention or Department of Health & Human Services.

**KEYWORDS:** *Rickettsia*, *Orientia*, genome, variable number of tandem repeats (VNTR), annotation

## ABSTRACT

The complete genome sequences available for eight species of *Rickettsia* and information for other near relatives in the Rickettsiales including *Orientia* and species of Anaplasmataceae are a rich resource for comparative analyses of the evolution of these obligately intracellular bacteria. Differences in these organisms have permitted them to colonize varied intracellular compartments, arthropod vectors and vertebrate reservoirs in both pathogenic and symbiotic relationships. We summarize some comparative aspects of the genomes of these organisms with particular attention to the recently completed sequence for *R. canadensis* McKiel strain and an estimated 2/3 of the genome sequence for a Thailand patient isolate of *Orientia tsutsugamushi*. The *Rickettsia* genomes exhibit a high degree of synteny punctuated by distinctive chromosome inversions and consistent phylogenetic relationships regardless whether protein coding sequences or RNA genes, concatenated open reading frames or gene regions or whole genomes are used to construct phylogenetic trees. The aggregate characteristics (number, length, composition, repeat identity) of tandem repeat sequences of *Rickettsia* which often exhibit recent and rapid divergence between closely related strains and species of bacteria are also very conserved in *Rickettsia* but differed significantly in *Orientia*. *O. tsutsugamushi* shared no significant synteny to species of *Rickettsia* or Anaplasmataceae supporting its placement in a unique genus. Like *Rickettsia felis*, *Orientia* has many transposases and ankyrin and tetratricopeptide repeat domains. *Orientia* shares the important ATP/ADP translocases and proline-betaine transporters multigene families with *Rickettsia* but has more gene families that may be involved in regulatory and transporter responses to environmental stimuli.

## INTRODUCTION

Since the first complete genome sequence of *Rickettsia prowazekii* Madrid E was obtained in 1998 and a close evolutionary relationship of rickettsiae to the ancestor of the mitochondria was proposed,<sup>1</sup> seven more genomes of *Rickettsia* have been sequenced. These include *Rickettsia conorii* strain Malish 7,<sup>2-3</sup> *Rickettsia typhi* Wilmington,<sup>4</sup> *Rickettsia felis* URRWCCal2,<sup>5</sup> *Rickettsia rickettsii* Sheila Smith CWPP (GenBank AADJ01000001), *Rickettsia sibirica* 246 (GenBank AABW000000001),<sup>6</sup>

*Rickettsia akari* Hartford CWPP (GenBankAAFE01000001), and *Rickettsia canadensis* McKiel (present authors). The genomic sequences of *Rickettsia bellii* and *Orientia tsutsugamushi* are also being obtained by several groups including the present authors and it is likely that genome sequences of other species and strains in the Order Rickettsiales will be soon available because of the small genome sizes of these species, the interesting diversity of their microbial, cell-biological and ecological traits, the continued discovery of new rickettsial agents, and their importance for human and veterinary medicine. This additional genomic sequence information provides a vast amount of information for analyzing the evolution and functional organization of the genomes of intracellular organisms during their adaptation to different intracellular niches and hosts. In this paper, we summarize some comparative aspects of the genome sequences of the seven species of *Rickettsia* and first describe some attributes of the genomes of *R. canadensis* and *Orientia tsutsugamushi*.

## **PROPERTIES AND PHYLOGENY OF *RICKETTSIA CANADENSIS* AND *ORIENTIA TSUTSUGAMUSHI***

McKiel *et al.*<sup>7</sup> first described isolation of a novel rickettsia, *R. canada* (sic) strain 2678. The agent was isolated in embryonated chicken eggs from a pool of *Haemaphysalis leporispalustris* ticks collected from a domestic rabbit used as a sentinel animal in field studies of tick transmitted diseases in snowshoe hares, *Lepus americanus*, and other small animals in the Richmond area near Ottawa, Ontario, Canada in 1962-63. A second isolate, strain H299, was recovered from a pool of 16 engorged ticks removed from a snowshoe hare shot in October 1964 about 20 miles south of the area where sentinel animals were stationed but this strain was lost. Inoculation of strain 2678 into guinea pigs, mice, hamsters, and rabbits caused seroconversion of these animals and raised antibodies that cross-reacted with *R. typhi* and *R. prowazekii* in complement fixation tests but toxin neutralization tests differentiated it from both *R. prowazekii* and *R. typhi*. Subsequent field work failed to detect this agent until another isolate CA410 of *R. canada* was obtained from *H. leporispalustris* tick from a black-tailed jack rabbit, *Lepus californicus* from Mendocino County in California.<sup>8</sup> Strains 2678 and CA410 differ somewhat in their protein profiles by SDS-polyacrylamide gel electrophoresis (PAGE) and in the reactivity and molecular

weight of their rOmpA proteins with monoclonal antibodies as detected by Western blotting<sup>9</sup> but only strain 2678 has been studied in much detail. Serologic evidence has been described that suggests *R. canadensis* may cause a febrile disease in man.<sup>10</sup>

The biological and phylogenetic placement of *R. canadensis* has been problematic. Some of its properties are summarized in TABLE 1. While *R. canadensis* is antigenically more closely related to typhus group rickettsiae, like pathogenic spotted fever group rickettsiae, it causes a generalized infection in ticks, can be transmitted by ticks and is maintained transstadially and transovarially in them;<sup>11,12</sup> it also invades the nuclei of cells<sup>11</sup> and has both rOmpA and rOmpB proteins like spotted fever rickettsiae.<sup>9,13</sup> On the other hand, like typhus group rickettsiae *R. canadensis* does not rapidly kill embryonated chicken eggs where it grows to high titers, it hemolyzes red blood cells, it is susceptible to erythromycin, and it forms small plaques more slowly than spotted fever group rickettsiae;<sup>14</sup> it also has a G+C % similar to that of typhus group rickettsiae.<sup>15</sup> By reassociation kinetics measurements, its genome size was estimated to be 35% bigger than those of *R. typhi* and *R. prowazekii* and 15% bigger than that of *R. rickettsii*.<sup>9</sup> *R. canadensis* is readily differentiated from *R. prowazekii* and *R. typhi* by specific activities of the enzymes glutamate dehydrogenase and glutamate oxaloacetate transaminase, its whole cell protein profiles by SDS PAGE, and isoelectric focusing patterns of soluble proteins.<sup>16</sup> DNA-DNA hybridization<sup>9</sup> and sequence comparisons of specific rickettsial genes like those of 16S rRNA and *groEL* (FIGURE 1) place it at a similar distance from typhus and spotted fever group rickettsiae much like the mite and flea transmitted species, *R. akari* and *R. felis*.<sup>17-18</sup> Similar comparisons with partial sequences of the 17 kDa antigen, citrate synthase, and 23S rRNA genes have been made.<sup>19-21</sup> This data suggested that *R. canadensis* may be representative of a distinct lineage of rickettsiae which is poorly represented as yet among characterized isolates but which, like *R. bellii* shares some ancestral characteristics with both classic typhus and spotted fever group rickettsiae; in that sense it may more closely resemble forms ancestral to this genus. Consequently, we obtained the complete genome sequence of the McKiel 2678 strain of *R. canadensis* and present here our initial observations on some of its genomic properties.

*Orientia tsutsugamushi* is the etiological agent of scrub typhus which is transmitted by larvae of numerous species of trombiculid mites. Although it has long been associated with the genus *Rickettsia* because of its similar obligate intracellular growth in the cytoplasm of infected host cells, in 1995 it was reclassified from *Rickettsia* to the genus *Orientia* due to its clear phenotypic and genotypic differences from that genus.<sup>22</sup> These traits included its different cell wall structure and apparent lack of peptidoglycan and lipopolysaccharide, also characteristics of the phagosome enclosed intracellular species in the genera *Anaplasma* and *Ehrlichia* but not *Rickettsia* which has both molecules.<sup>23</sup> *Orientia* also has a unique profile of antigens and proteins which distinguish it from all the other Rickettsiales.<sup>23-24</sup> Despite its monospecific status and in contrast to species of *Rickettsia*, *O. tsutsugamushi* has long been recognized to exhibit substantial antigenic diversity.<sup>25-26</sup> The widespread application of restriction fragment length polymorphism (RFLP) analysis and DNA sequencing to PCR amplicons obtained from isolates and from infected patient, rodent and chigger samples<sup>27-29</sup> has confirmed that the significant genetic diversity found in this species<sup>28-29</sup> is not confined to the type-specific surface antigen gene which is predominantly detected in serotyping.<sup>22,26,29</sup>

The 16S rDNA sequences of several prototype strains of *Orientia* formed a cluster with levels of similarity of greater than 98.5%<sup>30</sup> suggesting that forming new species in this genus is as yet unwarranted. However, this cluster clearly separated *O. tsutsugamushi* from the genus *Rickettsia* (90.2-90.6% similarity) and from the Anaplasmataceae (FIGURE 1). Interestingly, unlike the 16S rDNA tree, the phylogenetic tree with groEL gene<sup>31</sup> grouped *Orientia* more closely to *Anaplasma* and *Ehrlichia* than *Rickettsia* but with a lower bootstrap value than found with the 16S rDNA sequences (FIGURE 1). The G+C% composition of *Orientia* is 28.1-30.5%,<sup>22</sup> a value slightly lower than that found for typhus rickettsiae. The genome size of *Orientia* by pulsed field gel electrophoresis analysis has been reported to be between 2.4 and 2.7 Mbp,<sup>32</sup> about twice that of *Rickettsia* (TABLE 2). Currently, efforts to sequence strains of *Orientia* from Japan (Gilliam-like), Korea (Boryong), Papua/New Guinea (Karp), and Thailand (Northeast border) are underway in other laboratories. In order to assure that some of the genetic diversity found in different isolates of *Orientia* can be analyzed, and because the isolates available from

Thailand include many types not reported from other countries,<sup>25,28</sup> we chose to sequence the well-characterized human patient isolate AF PL-12 from Central Thailand.<sup>33</sup> Because this genome is larger than those of *Rickettsia* and it exhibits some of the complexity found in the slightly smaller *Bartonella henselae* genome as it also contains many repeated regions that were not amenable to completion by direct shotgun sequencing analysis,<sup>34</sup> it may require some concerted effort to complete the *Orientia* genome sequences. Consequently, we report here some of our observations on the estimated 70% (1.57 Mb) of the full genome sequence which has been obtained.

### **GENERAL ATTRIBUTES OF THE GENOME SEQUENCES OF *R. CANADENSIS* AND SEVEN OTHER SPECIES OF *RICKETTSIA***

Surprisingly, the assembled genome size of 1,159,772 bp for *R. canadensis* was much smaller than that obtained by reassociation kinetics<sup>15</sup> and only about 48 Kb bigger than those of *R. typhi* and *R. prowazekii* (TABLE 2). Its estimated coding capacity based on Glimmer predicted open reading frames was approximately 75.2% and its G+C% content 29% as expected. Like the other species of *Rickettsia* whose genomes have been sequenced, the genome of *R. canadensis* encodes 33 tRNA genes, the three (5S, 16S, and 23S) ribosomal RNA genes and three other RNA molecules, *ssrA* (tmRNA), *rnpB* (RNA component of ribonuclease P), and *srp* (signal recognition particle 4.5S RNA). Predicted open reading frames (ORF) based completely on automated GenMark and Glimmer analysis are generally close to the final results obtained by manual annotation for the four sequences that have complete annotations at NCBI (TABLE 2). Not surprisingly, assuming complete accuracy in sequencing, this is about ½ of the potential genes with greater than 50 aa size whether restricting the calls to ATG starts or including alternative start codons. This is because GenMark and Glimmer allow overlapping genes on the same strand but rarely include open reading frames that overlap in opposite orientation while Getorf is not restricted in this fashion and many small fragments overlap in the opposite orientation with larger ORFs. Although the number of ORFs predicted by Glimmer and GenMark might be expected to increase relative to *R. prowazekii* and *R. typhi* in direct proportion to their increased genome size, *R. canadensis* (4.3% bigger genome, 9.5-37.3% increase in ORFs), the four spotted fever group rickettsiae (10.8-14.2% bigger

genomes, 13.6-67.5% increase in ORFs) and *R. felis* (33.7 % bigger genome, 61-76 % increase in ORFs) have increasingly more predicted ORFs than expected for their respective % increases in genome sizes. Many of these predicted additional ORFs are small hypothetical “unique” proteins without homologues among these species or to other bacteria. This contrasts with the strong chromosome synteny and high gene homology of most well-identified functional proteins as described below; consequently, the reliability of ORF identification appears to decrease with increased genome size and many of these predicted ORFs may not be expressed or be essential.

PipMaker<sup>35</sup> was used to compare the genome sequence of *R. canadensis* to those of other species of *Rickettsia* (FIGURE 2). As is evident from the full genome strong diagonal lines found on these three plots and all the other pairwise comparisons (not shown), strong conservation of DNA sequences and specific gene orders is evident at both a local operon and genome-wide level among all eight of the genome sequences. This indicates that strong selective pressures have occurred in the genus to preserve the remarkable synteny observed among all of the species of *Rickettsia*. As is true for many other closely related bacteria, a cluster of relative inversions (upper left bottom right diagonals) has occurred in many of the species opposite the origin of replication (middle of PipPlots) but other parts of these genomes also have inversions less frequently. The presence of a unique large inversion encompassing the middle third of the chromosome was identified in *R. canadensis*. Inversion sites often occur close to tRNA genes. The presence of a significant level of off-diagonal signal is seen in all PipPlot pairwise comparisons including two of the four spotted fever group rickettsiae or three of these organisms (excluding *R. felis*) with *R. canadensis* (FIGURE 2). However, the amount of the off diagonal signal is greatly diminished when comparison is made to the genome sequences of *R. typhi* or *R. prowazekii* (FIGURE 2) or between *R. canadensis* and *R. felis* (not shown). This signal appears to be due to conservation of numerous small repetitive DNA elements scattered throughout rickettsial genomes. *R. canadensis* apparently shares more of these elements with spotted fever group rickettsiae than with *R. prowazekii* or *R. typhi*.

#### **TANDEM REPEAT SEQUENCES IN *RICKETTSIA* AND *ORIENTIA***

A wide variety of repetitive sequence elements are found in bacteria. These range from duplicated genes which exhibit varying degrees of homology like the ATP/ADP translocase genes, proline-betaine transporters, and *sca* gene families in *Rickettsia*, to the small palindromic repeat elements which are found in both coding and non-coding DNA, and direct large repetitive protein domains found in varying number in the rOmpA proteins. For the eight *Rickettsia* genomes and partial *Orientia* sequence we have examined in some detail the properties of repeated sequences (VNTR-variable number of tandem repeats) that vary in their copy number, composition, period length, and the amount and type of homology present between individual repeat copies (FIGURE 3). The prevalence and complexity of VNTR vary widely among different species of bacteria. VNTR are thought to be fast evolving loci because the number of repeats is commonly affected by slippage during replication. VNTR loci can be used to fingerprint differences between closely related strains and are therefore powerful tools for forensic science and molecular epidemiology.

A tandem repeat finder program was used to identify VNTR sites in all eight species of *Rickettsia* and in the partial sequence of *Orientia* available (FIGURE 3).<sup>36</sup> The number of repeats was fairly similar among the species of *Rickettsia* and differences did not correlate well with genome size or closest relative since *R. sibirica* (1.25 Mb) had the fewest repeats (347) while *R. typhi* (1.11 Mb) with 459 approached that of *R. felis* (1.49 Mb) which had 493 (FIGURE 3A). However, the *Orientia* sequence (1.57 Mb), which was only slightly greater than that of *R. felis*, had more than 200 more repeats; It also differed markedly in the number of larger repeats than those found in *Rickettsia*. In species of both genera the vast majority of repeats had a very low copy number between 2 and 3 and the distribution of repeats with copy number was very similar. Not surprisingly, given the low G+C % composition of the total DNAs of these 9 species, the VNTR also had a decided bias toward a low G+C% (FIGURE 3B). However, the distributions of VNTRs for all species were greatly skewed toward values less than the average genome compositions. This is somewhat surprising as typically 2/3 of the VNTR are located within coding sequences which are somewhat higher in G+C% than intergenic regions. However, factors contributing to increases in the numbers of these low G+C% VNTRs may have also influenced the accumulation of A



and T's in these genomes that lowered their overall compositions. *Orientia*, and to a lesser extent *R. felis*, were atypical in having a greater number of VNTRs with high G+C% than the other species.

The presence of exact repeats of sequences provides a possible site for unequal homologous recombination which may lead to loss or corresponding gain in the number of copies of a repeat between the two chromosomes exchanging the repeats. This may be responsible for alterations in the number of rOmpA 72 and 75 amino acid repeats in different strains of *R. conorii* and *R. rickettsii*.<sup>37-38</sup> On a smaller scale different strains of *R. rickettsii* have insertion/deletion (indel) events that may have occurred in this fashion<sup>39</sup> and differences in homologous genes often occur in the number of copies of tandem repeats found in different species.<sup>40</sup> Alterations in the repeats that make them imperfect repeats may be less subject to these events. Consequently, we examined the extent to which imperfect repeats were included in the VNTR regions of the eight *Rickettsia* genome sequences and for *Orientia* and then the number of repeats with indel events. In different species 38-50% of VNTRs exhibited complete identity and another 30-42% exhibited 90-98% repeat sequence identity so these constituted the majority of VNTRs (FIGURE 3C). Typically about 13-16% of the VNTR had 80-89% identity and only 2-4% had 70-79% identity. An even smaller number of indels were found in each of the species as 72-74% lacked indels and another 3-6% had only 1-5% of nucleotides as indels; 13-18% had 1-2 nucleotides as indels out of 20 in a VNTR (FIGURE 3D). Consequently, one would expect that variation in VNTR copy numbers could be a significant source of inter and intraspecies variation in both *Rickettsia* and *Orientia* and other than the total number of sites available for evolutionary change by various genetic processes, these organisms are very similar in their VNTR characteristics.

### **PHYLOGENETIC RELATIONSHIPS AMONG SPECIES OF *RICKETTSIA* AND *ORIENTIA***

As has been discussed above and shown in FIGURE 1, various genes have been used to examine phylogenetic relationships among different rickettsiae. Most commonly these have included 16S rDNA, *groEL*, 17 kDa and 120 kDa cytoplasmic antigen genes, rOmpA and rOmpB species specific antigens, *rpoB* and *gltA*. Some of these genes exhibit relatively little variation (17 kDa and 16S rDNA) while others like the 120 kDa, rOmpB and *rpoB* genes have evolved more rapidly. In principle this type of

analysis can be done for every gene which is present in all of the species of the genus (a limitation for rOmpA) or it is possible to concatenate a series of genes before tree construction to obtain an overall average for the species. In fact most of the individual gene trees have a very similar shape with occasional exchanges of positions with poorly supported bootstrap values. Because many of the type four secretion system proteins (T4SS) are very large and are thought to be essential for intracellular life by all the species in the Rickettsiales, they appear to be an excellent choice for construction of phylogenetic trees (FIGURE 4A). Indeed a tree constructed by examining the full genome sequences (FIGURE 4B) greatly resembled that for the T4SS proteins further supporting the validity of this approach. Because many T4SS proteins have now been identified in *Orientia* and in the Anaplasmataceae, we also constructed a phylogenetic tree for some of the VirB4 protein homologues (FIGURE 5). This tree clearly places *Orientia* more closely to *Rickettsia* than to Anaplasmataceae, a result similar to that obtained for the 16S rDNA tree (FIGURE 1). However, it also supports the taxonomic placement of *Orientia* as a unique genus.

### **SOME IMPORTANT ATTRIBUTES OF *ORIENTIA* FROM ITS PARTIAL ANNOTATION**

Although the full sequence of the *Orientia* genome is not yet available, it contained 27 tRNA genes, the 5S, 16S and 23S rDNA genes, DNA encoding signal recognition particle protein Ffh and its receptor FtsY but not signal recognition protein RNA, the RNA of ribonuclease P (*rnpB*) but not the protein gene (*rnpA*), and the tmRNA binding protein but not the tmRNA (*ssrA*). Only one of the 27 tRNA genes encoded a tRNA (Leu-CAA) that was not used by the species of *Rickettsia* but it is present in *Anaplasma marginale*, *Wolbachia* TRS and Dmel and *Ehrlichia ruminantium*. Homologues were also found for other genes previously described for *Orientia* such as the 56 kDa and 47 kDa antigen genes, superoxide dismutase, and a 3082 bp PCR target (GenBank L23765) but not the 22 kD antigen or groESL operon. The ADP/ATP translocase family of *Rickettsia* had three homologs (*tlc1*, 2 & 5) and the proline/betaine transporter family had 4 homologs in *Orientia*. The concatenated contigs obtained for the genome of *Orientia* were compared with the genome sequences of three species of Anaplasmataceae, *Anaplasma phagocytophilum*, *Neorickettsia sennetsu* and *Wolbachia pipientis* from *Drosophila*

*melanogaster* and to *R. canadensis* using PipMaker (FIGURE 6). No strong diagonal lines were observed in contrast to the results shown in FIGURE 2. Some off diagonal signal was obtained with all four genomes, but particularly with *R. canadensis*. Unsuccessful attempts to identify genes with BLAST using nucleotide searches of nucleotide databases confirmed the strong evolutionary divergence of *Orientia* from the others as shown in FIGURE 1 and 5. However, BLAST and SMART (Simple Modular Architecture Research Tool)<sup>37</sup> analyses of putative open reading frames identified using Glimmer and a new annotation tool called IMASS (developed by CS and GD) lead to the recognition of some significant attributes of the *Orientia* genome. First, a large number of repeated elements are present in the *Orientia* genome most notably including 19 SMART domains found in more than one ORF, most frequently the 56 histidine kinase-like ATPases domains (HATPase\_C) and the 37 domains found in ATPases associated with a variety of cellular activities (AAA) (TABLE 3). The number and type of SMART domains detected suggest that *Orientia* has a much larger range of regulatory proteins involved in responses to environmental stimuli than any known for *Rickettsia* including *R. felis*.<sup>6</sup> *R. felis* is also unusual among species of *Rickettsia* in having two plasmids, 82 transposase genes, 8 phage related proteins, 22 ankyrin and 11 tetratricopeptide (TPR)-containing ORFs.<sup>6</sup> In *Orientia* 82 transposase domains of 6 types have been detected (41 copies of transposase 11, 21 of transposase 14, 9 of transposase 31, 5 each of transposases 9 and 20, and 1 of transposase 8). The 42 ankyrin domain proteins found in *Orientia* varied in the numbers of repeats present (1-9), their lengths (28-56), sequences, and repeat patterns, and the sizes of their ORFs (57-938). TPR-containing ORFs were also common in *Orientia* and although their repeat length was generally 34 amino acids, the number of repeats and sizes of the ORFs containing TPR varied considerably. Given the many common elements found in *R. felis* which has two plasmids and the numerous plasmid and conjugation related proteins found in *O. tsutsugamushi*, it would not be very surprising if some isolates of this species also have plasmids and may be capable of conjugation.

### ACKNOWLEDGEMENTS

The studies reported in this article were supported by NIAID Grants AI05326-01 and 50942. The authors thank our international fellows Truong Thang and Jui-Shan Ma and our student interns Collins

Rainey, Holly Lemon-Steiner, Shwante Rogers, Erin Humbles, Noe Galvan, Albert Garcia, Cheryl Chow, Sasha Yurgionas, Drew MacKellar, Nina Tioleco, Matt Erdman, and Sara Scherer for their assistance in the annotations of the *Rickettsia* and *Orientia* genome sequences.

## REFERENCES

1. ANDERSSON, S. G., A. ZOMORODIPOUR, J. O. ANDERSSON, *et al.* 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. *Nature* **396**: 133-140.
2. OGATA, H., S. AUDIC, V. BARBE *et al.* 2000. Selfish DNA in protein-coding genes of *Rickettsia*. *Science* **290**: 347-350.
3. OGATA, H., S. AUDIC, P. RENESTO-AUDIFFREN, *et al.* 2001. Mechanisms of evolution in *Rickettsia conorii* and *R. prowazekii*. *Science* **293**: 2093-2098.
4. MCLEOD, M. P., X. QIN, S. E. KARPATHY, *et al.* 2004. Complete genome sequence of *Rickettsia typhi* and comparison with sequences of other rickettsiae. *J. Bacteriol.* **186**: 5842-5855.
5. MALEK, J. A., J. M. WIERZBOWSKI, W. TAO, *et al.* 2004. Protein interaction mapping on a functional shotgun sequence of *Rickettsia sibirica*. *Nucleic Acids Res.* **32**: 1059-1064.
6. OGATA, H., P. RENESTO, S. AUDIC, *et al.* 2005 The Genome Sequence of *Rickettsia felis* identifies the first putative conjugative plasmid in an obligate intracellular parasite. *PLoS Biol* **3**(8): e248.
7. MCKIEL, J. A., E. J. BELL, & D. B. LACKMAN. 1967. *Rickettsia canada*: a new member of the typhus group of rickettsiae isolated from *Haemaphysalis leporispalustris* ticks in Canada. *Can. J. Microbiol.* **13**:503-510.
8. LANE, R. S., R. N. PHILIP, & E. A. CASPER. 1981. Ecology of tick-borne agents in California. II. Further observations on rickettsiae. *In* *Rickettsiae and Rickettsial Diseases*. W. Burgdorfer & R. L. Anacker Eds.: 575-584. Academic Press, New York.
9. CHING, W.-M., G. A. DASCH, M. CARL, & M. E. DOBSON. 1990. Structural analyses of the 120-kDa serotype protein antigens of typhus group rickettsiae. Comparison with other S-layer proteins. *Ann. N. Y. Acad. Sci.* **590**: 334-351.

10. BOZEMAN, F. M., B. L. ELISBERG, J. W. HUMPHRIES, *et al.* 1970. Serologic evidence of *Rickettsia canada* infection of man. J. Infect. Dis. **121**: 367-371.
11. BURGDORFER, W. 1968. Observations on *Rickettsia canada*, a recently described member of the typhus group rickettsiae. J. Hyg. Epid. Microbiol. Immunol. **12**: 26-31.
12. BRINTON, L. P., & W. BURGDORFER. 1971. Fine structure of *Rickettsia canada* in tissues of *Dermacentor andersoni* Stiles. J. Bacteriol. **105**: 1149-1159.
13. DASCH, G. A., & A. L. BOURGEOIS. 1981. Antigens of the typhus group of rickettsiae: importance of the species-specific surface protein antigens in eliciting immunity. In *Rickettsiae and Rickettsial Diseases*. W. Burgdorfer & R. L. Anacker Eds.: 61-70. Academic Press, New York.
14. WOODMAN, D. R., E. WEISS, G. A. DASCH, & F. M. BOZEMAN. 1977. Biological properties of *Rickettsia prowazekii* strains isolated from flying squirrels. Infect. Immun. **16**: 853-860.
15. MYERS, W. F., & C. L. WISSEMAN, Jr. 1981. The taxonomic relationship of *Rickettsia canada* to the typhus and spotted fever groups of the genus *Rickettsia*. In *Rickettsiae and Rickettsial Diseases*. W. Burgdorfer & R. L. Anacker Eds.: 313-325. Academic Press, New York.
16. DASCH, G. A., J. R. SAMMS, & E. WEISS. 1978. Biochemical characteristics of typhus group rickettsiae with special attention to the *Rickettsia prowazekii* strains isolated from flying squirrels. Infect. Immun. **19**: 676-685.
17. STOTHARD, D. R., & P. A. FUERST. 1995. Evolutionary analysis of the spotted fever and typhus groups of *Rickettsia* using 16S rRNA gene sequences. Syst. Appl. Microbiol. **18**: 52-61.
18. ROUX, V., & D. RAOULT. 1995. Phylogenetic analysis of the genus *Rickettsia* by 16S rDNA sequencing. Res. Microbiol. **146**: 385-396.
19. AZAD, A. F., J. B. SACCI, JR., W. M. NELSON, *et al.* 1992. Genetic characterization and transovarial transmission of a typhus-like rickettsia found in cat fleas. Proc. Natl. Acad. Sci. U. S. A. **89**: 43-46.

20. ROUX, V., E. RYDKINA, M. EREMEEVA, & 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.
21. ANDERSSON, S. G., D. R. STOTHARD, P. FUERST, & C. G. KURLAND. 1999. Molecular phylogeny and rearrangement of rRNA genes in *Rickettsia* species. *Mol. Biol. Evol.* **16**: 987-995.
22. TAMURA, A., N. OHASHI, H. URAKAMI, & S. MIYAMURA. 1995. Classification of *Rickettsia tsutsugamushi* in a new genus, *Orientia* gen. nov., as *Orientia tsutsugamushi* comb. nov. *Int. J. Syst. Bacteriol.* **45**: 589-591.
23. DASCH, G. A., E. WEISS, & J. C. WILLIAMS. 1990. Antigenic properties of the Ehrlichiae and other Rickettsiaceae. *In* Ehrlichiosis. A Vector-borne Disease of Animals and Humans. J. C. Williams & I. Kakokma Eds.:32-58. Kluwer Academic Publishers, Dordrecht, The Netherlands.
24. OAKS, E. V., R. M. RICE, D. J. KELLY, & C. K. STOVER. 1999. Antigenic and genetic relatedness of eight *Rickettsia tsutsugamushi* antigens. *Infect. Immun.* **57**: 249-253.
25. ELISBERG, B. L., J. M. CAMPBELL, & F. M. BOZEMAN. 1968. Antigenic diversity of *Rickettsia tsutsugamushi*: epidemiologic and ecologic significance. *J. Hyg. Epidemiol. Microbiol. Immunol.* **12**: 18-25.
26. MURATA, M., Y. YOSHIDA, M. OSONO, et al. 1986. Production and characterization of monoclonal strain-specific antibodies against prototype strains of *Rickettsia tsutsugamushi*. *Microbiol. Immunol.* **30**: 599-610.
27. KELLY, D., G. A. DASCH, T. C. CHAN, & T. M. HO. 1994. Detection and characterization of *Rickettsia tsutsugamushi* (Rickettsiales: Rickettsiaceae) in infected *Leptotrombidium* (*Leptotrombidium*) *fletcheri* chiggers (Acari: Trombiculidae) with the polymerase chain reaction. *J. Med. Entomol.* **31**: 691-699.
28. DASCH, G. A., D. STRICKMAN, G. WATT, & C. EAMSILA. 1996. Measuring genetic variability in *Orientia tsutsugamushi* by PCR/RFLP analysis: a new approach to questions about

- its epidemiology, evolution, and ecology. *In Rickettsiae and Rickettsial Diseases*. J. Kazar & R. Toman, Eds.: 79-84. VEDA, Bratislava, Slovak Republic.
29. TAMURA, A. 1999. Genetic diversity in *Orientia tsutsugamushi*. *In Rickettsiae and Rickettsial Diseases at the Turn of the Third Millenium*. D. Raoult & P. Brouqui, Eds.: 67-73. Elsevier, Paris, France.
  30. OHASHI, N., M. FUKUHARA, M. SHIMADA, & A. TAMURA. 1995. Phylogenetic position of *Rickettsia tsutsugamushi* and relationship among its antigenic variants by analyses of 16S rRNA gene sequences. *FEMS Microbiol. Lett.* **125**: 299-304.
  31. STOVER, C. K., D. MARANA, G. DASCH, & E. OAKS. 1990. Molecular cloning and sequence analysis of the Sta58 major antigen gene of *Rickettsia tsutsugamushi*: sequence homology and antigenic comparison of Sta58 to the 60-kilodalton family of stress proteins. *Infect. Immun.* **58**: 2076-2084.
  32. TAMURA, A. 1996. Overview of *Orientia tsutsugamushi* based on recent findings. *In Rickettsiae and Rickettsial Diseases*. J. Kazar & R. Toman, Eds.: 62-71. VEDA, Bratislava, Slovak Republic.
  33. STRICKMAN, D., C. D. SMITH, K. D. CORCORAN, *et al.* 1994. Pathology of *Rickettsia tsutsugamushi* infection in *Bandicota savilei*, a natural host in Thailand. *Am. J. Trop. Med. Hyg.* **51**: 416-423.
  34. ALSMARK, C. M., A. C. FRANK, E. O. Karlberg, *et al.* 2004. The louse-borne human pathogen *Bartonella quintana* is a genomic derivative of the zoonotic agent *Bartonella henselae*. *Proc. Natl. Acad. Sci. U.S.A.* **101**: 9716-9721.
  35. SCHWARTZ, S., Z. ZHANG, K. A. FRAZER, *et al.* 2000. PipMaker: a web server for aligning two genomic DNA sequences. *Genome Res.* **10**: 577-586.
  36. BENSON, G. 1999. Tandem repeats finder: a program to analyze DNA sequences. *Nucleic Acids Res.* **27**: 573-580.

37. WALKER, D. H., H. M. FENG, J. I. SAADA, *et al.* 1995. Comparative antigenic analysis of spotted fever group rickettsiae from Israel and other closely related organisms. *Am. J. Trop. Med. Hyg.* **52**: 569-576.
38. MATSUMOTO, M., Y. TANGE, T. OKADA, *et al.* 1996. Deletion in the 190 kDa antigen gene repeat region of *Rickettsia rickettsii*. *Microb. Pathog.* **20**: 57-62.
39. EREMEEVA, M. E., R. M. KLEMT, L. A. SANTUCCI-DOMOTOR, *et al.* 2003. Genetic analysis of isolates of *Rickettsia rickettsii* that differ in virulence. *Ann. N. Y. Acad. Sci.* **990**: 717-722.
40. GILMORE, R. D., JR. 1993. Comparison of the *rompA* gene repeat regions of rickettsiae reveals species-specific arrangements of individual repeating units. *Gene* **125**: 97-102.
41. LETUNIC, I., R. R. COPLEY, S. SCHMIDT, *et al.* 2004. SMART 4.0: towards genomic data integration. *Nucleic Acids Res.* **32**(Database issue): D142-D144.
42. KUMAR, S., K. TAMURA, I. B. JAKOBSEN, & M. NEI. 2001. MEGA2: Evolutionary genetics analysis software. *Bioinformatics* **17**: 1244-1245.
43. BRAY, N., & L. PACHTER. 2004. MAVID:constrained ancestral alignment of mutiple sequences. *Genome Res.* **14**: 693-699.
44. SHAW, C. D., M. E. EREMEEVA, H. H. SHAH, AND G. A. DASCH. 2005. Development of an Interactive Multigenome Analysis System (IMASS) for annotation of bacterial sequences. Abstract 52 (F). 2005 American Society for Microbiology Biodefense Research Meeting, Baltimore, MD.



## FIGURE LEGENDS

**FIGURE 1.** Phylogenetic relationships of selected Rickettsiales whose genomic sequences are available. Neighbor-joining phylogenetic trees based on 16S rRNA gene [top] and GroEL gene [bottom] alignments were drawn using the MEGA2 software.<sup>42</sup> Distance matrix was calculated using Kimura-2 parameters and based on 1317 sites for 16S rRNA gene and 1217 sites for GroEL. The scale bar represents % sequence divergence. Numbers at nodes are the percentage of 1,000 bootstrap resamplings that support the topology shown.

**FIGURE 2.** PipMaker plot (<http://pipmaker.bx.psu.edu/cgi-bin/pipmaker?basic>) comparison of genome sequences of *R. canadensis* with genome sequences of three other species of *Rickettsia*.

**FIGURE 3.** Comparative analysis of tandem repeat sequences found in genome sequences of eight *Rickettsia* and *Orientia*. **A.** Percentage of repeats with different copy numbers (total number of repeats given beside each species name). **B.** Distribution by G+C% composition of number of tandem repeat sequences. **C.** Distribution of number of repeats by percentage of sequence homology to the individual repeat unit. **D.** Prevalence of percentage of indel nucleotides in the tandem repeat sequences.

**FIGURE 4 A.** Phylogenetic analysis of 14 concatenated T4SS complex protein sequences (8032 amino acid sites of homologs of RT0034, RP103-RP108, RP286-RP291, RP294, and RP784) of *Rickettsia* constructed as in FIGURE 1. **B** *Rickettsia* phylogeny based on whole genome sequence comparisons using Mavid.<sup>43</sup>

**FIGURE 5.** Phylogenetic relationship VirB4 amino acid sequences of species of Rickettsiales. Neighbor-joining phylogenetic tree based on VirB4 amino acid sequence similarities (738 sites) was drawn using the MEGA2 software as described in FIGURE 1. NCBI entrez genome identifiers are indicated for ORFs of species with paralogous sequences used in the tree.

**FIGURE 6.** PipMaker Plot comparison of the genome sequences of *Orientia* with four other species of Rickettsiales.

**TABLE 1. Biological and molecular characteristics of *R. canadensis* shared with either typhus or spotted fever group rickettsiae**

Characteristic	Typhus group	Spotted fever group
Tick association		++
Growth in yolk sac	++	
Plaque formation	++	
Intracellular growth		++
Hemolytic activity	++	
S-layer		rOmpA + rOmpB
Lipopolysaccharide type	++	
Cross-protective activity	++	
Genome size		1.33 x TG <sup>1</sup>
G+C%	29.2	

<sup>1</sup>TG, typhus group; the values are obtained based upon DNA reassociation method.<sup>15</sup> The exact numbers are not shown since it appears that the original publication appears to have a systematic error in estimation of rickettsial genome size.

**TABLE 2. Putative numbers of open reading frames (ORF) found in *Rickettsia* chromosomes using different recognition algorithms<sup>1</sup>**

<b>Species Total bp</b>	<b>GenMark</b>	<b>Glimmer</b>	<b>Getorf ATG only (&gt;50 aa)</b>	<b>Getorf w. Alt. starts (&gt;50 aa)</b>	<b>NCBI Annotation</b>
<i>R. canadensis</i> 1,159,772	1,032	1,177	1,505	2,267	MAI <sup>2</sup>
<i>R. felis</i> 1,485,148	1512	1510	2231	3263	1400
<i>R. akari</i> 1,231,204	1,337	1,285	1,862	2,786	MAI
<i>R. conorii</i> 1,268,755	1,404	1,532	1,962	2,862	1374
<i>R. rickettsii</i> 1,257,710	1,365	1,527	1,914	2,849	MAI
<i>R. sibirica</i> 1,250,021	1,443	1,357	1,893	2,809	MAI
<i>R. prowazekii</i> 1,111,523	903	982	1,328	1,889	835
<i>R. typhi</i> 1,111,496	860	854	1,319	1,879	877

<sup>1</sup>Details on ORF recognition algorithms can be found at the following websites:

<http://opal.biology.gatech.edu/GeneMark/> (GenMark),

<http://www.genomics.jhu.edu/Glimmer/> (Glimmer), and <http://emboss.sourceforge.net/> (getorf).

<sup>2</sup>Manual annotation incomplete

**TABLE 3. Type and prevalence of 223 SMART domains<sup>1</sup> found in the partial genome sequence of *Orientia***

SMART Domains Detected	No. Found
HATPase_c (Histidine kinase-like ATPases)	56
AAA (ATPases associated with a variety of cellular activities)	37
HDc (Metal dependent phosphohydrolases with conserved 'HD' motif)	23
HhH1 (Helix-hairpin-helix DNA-binding motif class 1)	12
KH (K homology RNA-binding domain)	10
REC (cheY-homologous receiver domain)	7
S1 (Ribosomal protein S1-like RNA-binding domain)	6
HisKA (His Kinase A phosphoacceptor domain)	5
Elp3 (Elongator protein 3, MiaB family, Radical SAM)	4
CBS, HNHc, HTH_XRE, PHB, TOPRIM	3 each
DNAJ, KOW, PDZ, PlsC, S4	2 each
53EXOc, Ami_2, BHL, BRCT, CSP, DESDc, DSRM, DUF1, DUF2, ENDO3c, FES, GuKc, HELICc, HhH2, NHN_nucleases, LIGANc, LON, MUTSac, MUTSd, NGN, PAC, ParB, PBPb, PLDc, POLAc, rADc, RHOD, RIBOc, RL11, RPOLA_N, RPOLD, small_GTPase, STYKc, TOP1Ac, TOP1Bc, TOP2c, YqgFc, ZnF_C2HC, ZnF_CHCC	1 each

<sup>1</sup><http://smart.embl.de>

**TABLE 4. High quality identification of some *Orientia* genes by BLAST**

<b>Conserved Genes</b>	<b>High Interest Genes</b>
endonuclease III ( <i>nth</i> )	2-acylglycerophosphoethanolamine acyltransferase-2 ( <i>aas</i> )
ATP-dependent nuclease subunit A ( <i>addA</i> )	3-octaprenyl-4-hydroxybenzoate carboxy-lyase ( <i>ubiX</i> )
NAD-dependent DNA ligase ( <i>ligA</i> )	ADP, ATP carrier protein ( <i>tlc2</i> )
excinuclease A subunit ( <i>uvrA</i> )	hemolysin C ( <i>tlyC</i> )
DNA mismatch repair protein MutS ( <i>mutS</i> )	GTP-binding protein Era ( <i>era</i> )
elongation factor P ( <i>efp</i> )	heme exporter protein C ( <i>ccmC</i> )
histidyl tRNA synthetase ( <i>hisS</i> )	nitrogen assimilation protein ( <i>ntrX</i> )
isoleucyl-tRNA synthetase- ( <i>ileS</i> )	nitrogen regulation protein ( <i>ntrY</i> )
lysyl-tRNA synthetase ( <i>lysS</i> )	proton-sodium glutamate symport protein ( <i>gltP</i> )
poly(A) polymerase ( <i>pcnB</i> )	outer membrane assembly protein ( <i>asmA</i> )
replicative DNA helicase ( <i>dnaB</i> )	signal peptidase I ( <i>lepB</i> )
ribonuclease III ( <i>rnc</i> )	dihydrodipicolinate synthetase ( <i>dapA</i> )
succinyl CoA synthetase, alpha subunit ( <i>sucD</i> )	integrase/recombinase ( <i>xerD</i> )

