Molecular Phylogeny and Rearrangement of rRNA Genes in *Rickettsia* Species

Siv G. E. Andersson, Diane R. Stothard,†[1] Paul Fuerst,† and Charles G. Kurland**

*Department of Molecular Evolution, University of Uppsala, Uppsala, Sweden; and †Department of Molecular Genetics, Ohio State University

It has previously been observed that *Rickettsia prowazekii* has an unusual arrangement of the rRNA genes. In this species, the three rRNA genes, 16S (*rrs*), 23S (*rrl*), and 5S (*rrf*), are not linked in the typical arrangements for bacteria. Rather, the 16S rRNA gene has been separated from the 23S and 5S rRNA gene cluster, and the 23S rRNA gene is preceded by a gene which codes for methionyl-$tRNA_f^{Met}$ formyltransferase (*fmt*). In this study, we screened the genus *Rickettsia* for the *fmt-rrl* motif in order to examine the phylogenetic depth of this unusual rRNA gene organization. A rearranged operon structure was observed in *Rickettsia conorii, Rickettsia parkeri, Rickettsia sibirica, Rickettsia rickettsii, Rickettsia amblyomii, Rickettsia montana, Rickettsia rhipicephali, Rickettsia australis, Rickettsia akari, Rickettsia felis, Rickettsia canada,* and *Rickettsia typhi.* There is also evidence for a divided operon in *Rickettsia belli,* but in this species, the *fmt* gene could not be identified upstream of the 23S rRNA gene. In order to place the rearrangement event in the evolutionary history of the *Rickettsia,* phylogenetic analyses were performed based on the *fmt-rrl* spacer regions and the 23S rRNA genes. Based on these phylogenies, we suggest that the genomic rearrangement of the rRNA genes preceded the divergence of the typhus group and the spotted fever group *Rickettsia.* The unique organization of the 23S rRNA genes provides a simple diagnostic tool for identification of *Rickettsia* species.

Introduction

Rickettsiae are obligate intracellular, gram-negative bacteria that have natural arthropod hosts: ticks, mites, and insects (Weiss and Moulder 1984). They are capable of infecting humans and other vertebrates, but they are frequently pathogenic in these secondary hosts. They are phylogenetically classified with the α-Proteobacteria (Weisburg et al. 1989; Olsen, Woese, and Overbeek 1994). More specifically, they are members of the family Rickettsiaceae, which includes the three tribes Rickettsieae, Ehrlichieae, and Wolbachieae (Weiss and Moulder 1984). The genus *Rickettsia* has historically been divided on the basis of immunological cross-reactivity and vector organism into three biotypes: the spotted fever group (SFG), the typhus group (TG), and the scrub typhus group (STG). However, the STG species, *Rickettsia tsutsugamushi,* has recently been reclassified into a novel genus, *Orientia* (Tamura et al. 1995).

Members of the SFG are distributed worldwide. All are established in hard body ticks, *Ixodidae,* with the exception of *Rickettsia akari,* which is transmitted by a mesostigmatic mite. These bacteria grow and divide primarily within the host cell cytoplasm but are capable of multiplying within the cell nucleus as well (Wisseman et al. 1976). They cause disease in humans such as Rocky Mountain spotted fever (*Rickettsia rickettsii*), Siberian tick typhus (*Rickettsia sibirica*), rickettsial pox (*R. akari*), and Queensland tick typhus (*Rickettsia australis*). The TG is composed of two insect-borne species, *Rickettsia prowazekii* and *Rickettsia typhi,* and one tick-borne species, *Rickettsia canada* (Weiss and Moulder 1984). *Rickettsia prowazekii* and *R. typhi* grow and divide exclusively within the host cell cytoplasm and are the etiological agents of epidemic and murine typhus, respectively. *Rickettsia canada* exhibits no known human disease (McKeil, Bell, and Lackman 1967). The STG of *Rickettsia* is composed of several strains of a single species, *Orientia tsutsugamushi* (Tamura et al. 1995). It is transmitted by chiggers of trombiculid mites and is the causative agent of scrub typhus.

There are several species which do not fall distinctly into any of the three biotypes based on phenotypic data. *R. akari* is a member of the SFG based on immunological cross-reactivity and its acarine ecology. However, it also cross-reacts with sera prepared against TG species. *Rickettsia bellii* is capable of intranuclear growth and is tick-borne, but it cross-reacts with sera prepared against SFG and TG species (Anacker, Mann, and Gonzales 1987; Philip et al. 1983). However, sequence comparisons of the *R. bellii* 16S and 23S rRNA genes (Stothard, Clark, and Fuerst 1994; Stothard and Fuerst 1995) and the citrate synthase gene (Roux et al. 1977) with SFG and TG rickettsia suggest that *R. bellii* diverged prior to the separation of the SFG and TG lineages. *Rickettsia canada,* despite its acarine ecology and intranuclear growth (Burgdorfer and Brinton 1970), has been placed in the TG based on a strong reaction with sera prepared against TG rickettsia (Weiss and Moulder 1984). The ELB agent, *Rickettsia felis,* was first identified with the TG rickettsia based on the observation that it causes a disease similar to murine typhus in human populations (Azad et al. 1992; Williams et al. 1992). However, sequence comparisons of the 17-kDa antigen gene and the 16S rRNA gene have subsequently shown that this species may be more closely related to the SFG rickettsia than to the TG rickettsia (Azad et al. 1992; Williams et al. 1992).

[1] Present address: Department of Medicine, Indiana University, Indianapolis.

Key words: molecular evolution, phylogeny, *Rickettsia,* rRNA genes.

Address for correspondence and reprints: Siv G. E. Andersson, Department of Molecular Evolution, University of Uppsala, Box 590, BMC, S-751 24 Uppsala, Sweden. E-mail: siv.andersson@molbio.uu.se.

Mol. Biol. Evol. 16(7):987–995. 1999

Features common to all the different species of *Rickettsia* are their intracellular growth environment, their small genome sizes (1.1 Mb) (Eremeeva, Roux, and Raoult 1993), and their low genomic G+C contents (29%–33%). The complete genome sequence for one member of this genus, *R. prowazekii*, has recently been determined (Andersson et al. 1998). Information obtained from the genome sequence supports the idea that the transition to the intracellular growth environment has been associated with massive losses of genetic material and extensive genomic rearrangements (Andersson et al. 1998). We have previously suggested that the single-copy number of the rRNA genes and the disruption of the traditional rRNA operon structure in *R. prowazekii* are expected consequences of such a reductive mode of evolution (Andersson et al. 1995; Andersson and Kurland 1995). We have also shown that the 23S rRNA gene (*rrl*) is preceded by the methionyl-$tRNA_f^{Met}$ formyltransferase gene (*fmt*) in *R. prowazekii*. In this study, we examine an additional 14 *Rickettsia* species to measure the incidence of this unusual arrangement of rRNA genes. In order to place the disassociation event in the evolutionary history of *Rickettsia*, we have reconstructed a phylogeny based on the sequences of the 23S rRNA genes and the *fmt-rrl* intergenic regions.

Materials and Methods

Experimental Procedures

Genomic DNA was isolated as described previously (Pretzman et al. 1987) from the following species: *R. akari* strain Hartford, *R. amblyommii* strain MO 85-1084, *R. australis* strain Philips, *R. bellii* strain 369-C, *R. canada* strain McKeil, *R. conorii* strain ITT-586, *R. felis* strain ELB, *R. montana* strain OH 83-441, *R. parkeri* strain Maculatum-20, *R. prowazekii* strain 22-2, *R. rhipicephali* strain 3-7-6, *R. rickettsii* strain Sawtooth, *R. sibirica* strain 246, *R. typhi* strain Wilmington, and *O. tsutsugamushi* strain Gilliam.

The primary PCR amplifications of the spacer region between *fmt* and *rrl* were carried out using one primer (ot56+2) from within the *fmt* gene, (CTAAAGCAGAAGGAAAAATT) and one from within the 23S rRNA gene, (GCTAGGCCGTACCCGGTACG). Reamplifications were carried out using the primer ot56+2 in combination with the nested primer ot56-2 (GCTTCTAGTGCCAAGGCATC) from the beginning of the 23S rRNA gene, as previously described (Andersson et al. 1995). The primary PCR amplifications of the 23S rRNA genes were based on primers designed from known conserved regions in the rRNA genes of several Proteobacteria, as described previously (Stothard, Clark, and Fuerst 1994). Finally, species-specific primers were used to close the short sequence gap between the 3′ end of the spacer region and the 5′ end of the 23S rRNA gene.

PCR amplifications employed standard conditions using 150 ng of genomic DNA and 1 μM of the primers. Reamplifications were carried out using 1 μl of a 1:300 to 1:600 dilution of the primary PCR product. The annealing temperature was 53°C for PCR of the spacer regions and 48°C for the amplification of the 23S rRNA genes. The PCR products were purified directly from the PCR reaction mixture using the Wizard™ PCR Preps DNA Purification System (Promega).

The double-stranded PCR products were sequenced by an autocycle sequencing protocol with the Thermo Sequenase cycle sequencing kit (Amersham) using fluorescently labeled primers or with the double-stranded cycle sequencing kit (Life Technologies) using γ-^{32}P 5′-end-labeled sequencing primers. In each reaction, 0.5 μg of PCR product and 10–20 pmol of primers were used. The products of the fluorescently labeled sequencing reactions were separated and analyzed on an A.L.F. sequencer (Pharmacia, Biotech, Norden), and the γ-^{32}P-labeled products were separated on an 8% polyacrylamide gel and visualized by exposure to X-ray film (Amersham). To ensure sequence fidelity, sequences were obtained from both strands and compared with the secondary-structure model for the large-subunit rRNA gene (Gutell, Gray, and Schnare 1993). However, a total of 15 positions in the rRNA genes could not be unambiguously resolved. These are marked as R (A/G), Y (C/T) or N (A/G/C/T).

Phylogenetic Analysis

The *fmt* genes and the *fmt-rrl* spacer regions were aligned using CLUSTAL W (Thompson, Higgins, and Gibson 1994) and manually inspected for misalignments. Nonsynonymous substitutions (nucleotide substitutions that result in an amino acid change) and synonymous substitutions (nucleotide substitutions that do not result in an amino acid change) were calculated as K_A and K_S, respectively, using Li's (1993) method. The 23S rRNA sequences were aligned with the assistance of ESEE (Cabot and Beckenbach 1989) after identifying homologous positions on the basis of secondary structure (Gutell, Gray, and Schnare 1993). Pairwise substitution frequencies in the spacers and the 23S rRNA gene sequences were estimated using the interface program phylo_win (Galtier, Gouy, and Gautier 1996). Phylogenetic relationships between the different *Rickettsia* species were estimated by the maximum-parsimony and distance neighbor-joining methods (Saitou and Nei 1987) in the phylo_win package (Galtier, Gouy, and Gautier 1996). Distance matrices were calculated using the Kimura two-parameter model to correct for multiple substitutions. Bootstrap values for the trees were obtained from 500 randomly generated trees. Sequence data from the 23S rRNA gene of *O. tsutsugamushi* Gilliam were used as the outgroup in the analysis to place the root joining the *Rickettsia* species.

Nucleotide Sequence Accession Numbers

The partial *fmt* genes, the downstream spacer, and the 23S rRNA gene sequences have been deposited in the EMBL, GenBank, and DDBJ nucleotide sequence databases under accession numbers Y13122–Y13132. The 23S rRNA gene sequences for *R. australis* and *R. canada* have been deposited in the EMBL, GenBank, and DDBJ nucleotide sequence databases under accession numbers AJ133711 and AJ133712, respectively.

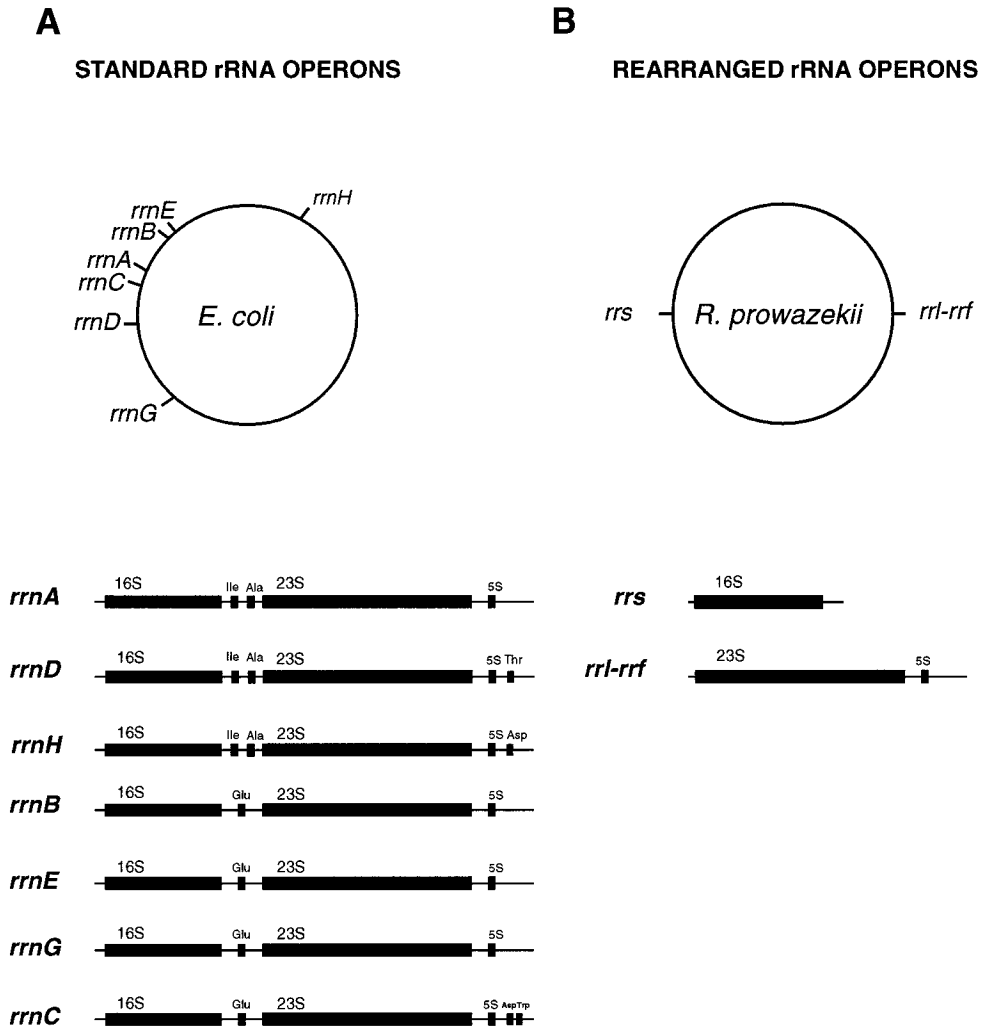

FIG. 1.—Schematic illustration of the number and organization of the rRNA genes in (*A*) *Escherichia coli* and (*B*) *Rickettsia prowazekii.*

The reported 23S rRNA gene sequences for *R. bellii* (U11015), *R. rickettsii* (U11022), and *R. prowazekii* (Z49077) were also used in the analysis.

Results

Rearrangement of the 23S rRNA Genes in *Rickettsia* Species

The rRNA genes of bacteria are normally organized into operons with the general structure 16S-23S-5S, and tRNA genes are often found in the spacer between the 16S and the 23S rRNA genes (Krawiec and Riley 1990). Multiple copies of the rRNA gene operons are frequently observed in free-living bacteria. For example, the *Escherichia coli* genome contains seven copies of the rRNA gene operons arranged in the standard manner (Blattner et al. 1997) (fig. 1*A*). In contrast, *R. prowazekii* has a single copy of each of the rRNA genes, and the 16S rRNA gene is located more than 500 kb away from the 23S-5S rRNA genes (Andersson et al. 1995, 1998) (fig. 1*B*). We have previously shown that the 23S-5S rRNA gene cluster (*rrl-rrf*) is preceded by the *fmt* gene in *R. prowazekii* (Andersson et al. 1995, 1998).

To study the distribution of this unusual organization of genes within the genus *Rickettsia*, we used PCR to amplify the spacer region between these two genes for 14 additional rickettsial species using one primer from within the *fmt* gene (ot56+2) and one primer from the beginning of the *rrl* gene (ot56−2) (fig. 2*A*). When genomic DNA from *R. prowazekii* was used to drive the PCR reaction, a product of 1.0 kb in size was obtained, as previously observed (Andersson et al. 1995). Successful PCR amplifications were also obtained when these primers were used in PCR amplifications together with genomic DNA from the following species: *R. akari, R. amblyommii, R. australis, R. canada, R. conorii, R. felis, R. montana, R. parkeri, R. rhipicephali, R. rickettsii, R. sibirica,* and *R. typhi.* Some size variation of the amplified PCR fragments was observed. Most notably, genomic DNA from *R. felis* generated a PCR product that was only 0.5 kb in length, as compared with ca. 1 kb for most other species. The identity of each PCR product was verified by sequence analysis. We concluded that the *fmt* and *rrl* genes are located near each other in each of these 13 rickettsial species (fig. 2*B*).

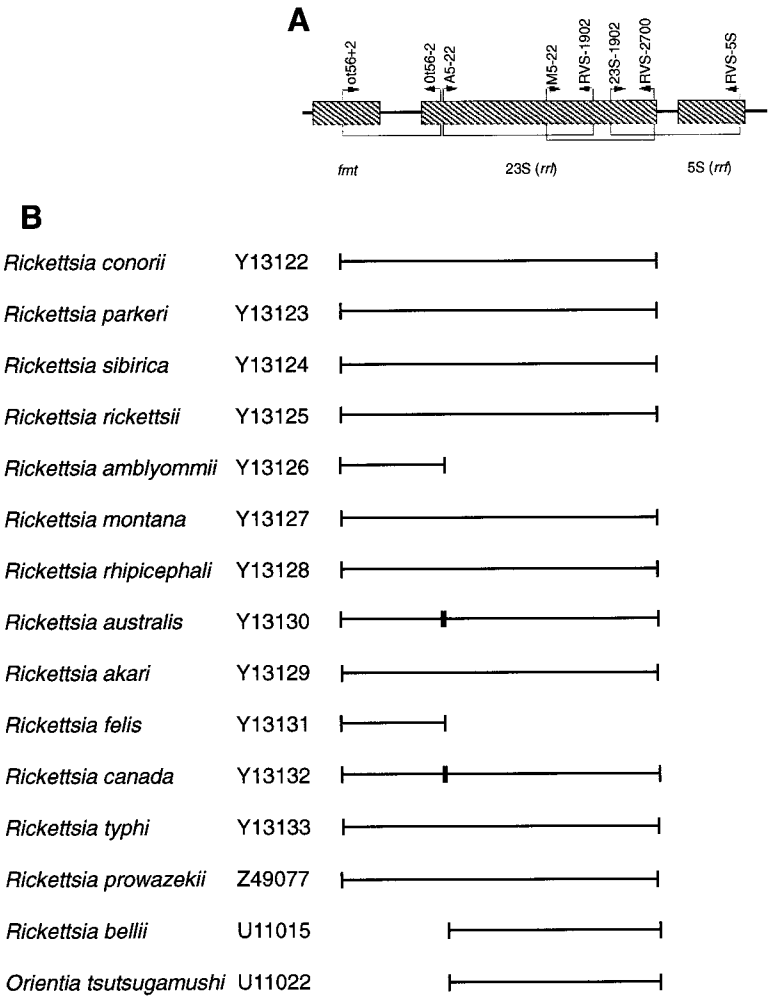

FIG. 2.—Rickettsial species studied. Schematic representation of (*A*) the organization of the 23S rRNA gene in *Rickettsia prowazekii* and primers used for PCR amplification of the *fmt-rrl* spacer region and the *rrl* gene and (*B*) the lengths and organizations of the sequences analyzed in this study. The numbers are accession numbers for corresponding sequences.

However, we were unable to obtain a PCR product when using genomic DNA from *R. bellii* or *O. tsutsugamushi* in combination with the primers designed for *fmt* and *rrl*. Previous attempts to amplify the 16S-23S rRNA intergenic spacer from *R. bellii* were unsuccessful using a variety of 16S and 23S rRNA primers (Stothard 1995). Clark (1990) produced a genomic clone of the 16S rRNA gene in *R. bellii* which contained the full-length 16S rRNA gene as well as 600 bp upstream and 900 bp downstream sequence. The downstream sequence of the 16S rRNA gene in this species did not contain sequence identifiable as *rrl* or any other long open reading frame. This suggests that the two rRNA genes are not linked to each other in *R. bellii*, as also supported by a Southern blot analysis (Clark 1990). However, the exact arrangements of the rRNA genes in *R. bellii* is currently unknown.

Sequence Features of the *fmt-rrl* Spacer Region

We observed a small but marked difference between members of the TG (*R. prowazekii* and *R. typhi*) and the classical SFG rickettsia (*R. conorii, R. sibirica, R. rickettsii, R. parkeri, R. amblyommii, R. montana,* and *R. rhipicephali*) with respect to both spacer lengths and nucleotide frequencies. Within the classical SFG rickettsia, the lengths of the spacer regions range from 845 to 888, and G+C content values range from 27.8% to 29.2%. The four most closely related members of this group, *R. conorii, R. sibirica, R. rickettsii,* and *R. parkeri,* have G+C content values that are almost identical (28.0%–28.6%). Within the TG rickettsia, the spacer sequences are 717 bp (*R. prowazekii*) and 733 bp (*R. typhi*) and have G+C content values of 22.5% and 23.6%, respectively. The G+C content value of the *fmt-rrl* spacer region in *R. prowazekii* is close to the average value of 21.4%, calculated from a large number of *R. prowazekii* intergenic regions (Andersson and Sharp 1996; Andersson et al. 1998). Taken together, the data show that the TG rickettsia have *fmt-rrl* spacer sequences that are more than 100 nt shorter in length and have G+C content values that are about 5% lower than the spacer sequences of the classical SFG rickettsia.

Within the so-called AAE clade (*R. australis, R. akari,* and *R. felis*) (Stothard 1995), G+C content values are intermediate between those associated with the TG and the SFG rickettsia (26.4%–27.9%). Spacer lengths

are similar to those for the SFG rickettsia (ca. 850 bp), with the exception of *R. felis*, which has a spacer sequence that is only 305 nt long. The reduced size of the spacer region of *R. felis* is the result of a contiguous deletion of almost 500 bp within this region. The deletion extends from the region immediately next to the *fmt* termination codon and an additional 500 bp into the spacer region. Since *R. canada* is believed to represent an early-branching group with uncertain affiliation (Stothard 1995; Roux et al. 1997), we have here treated it separately from the others. This species has a spacer length of 827 bp, which is intermediate between the values observed for *R. australis* and *R. akari* (852 and 858 bp) and for *R. prowazekii* and *R. typhi* (717 and 733 bp). The G+C content of 25.6% is also intermediate between those associated with members of the AAE clade (27%) and the TG rickettsia (23%).

Substitutions and Deletions Within the Spacer Regions

The *fmt-rrl* spacer sequences were aligned, and the substitution frequency for members of the TG and the SFG rickettsia was estimated to be, on average, 0.20 substitutions per site (table 1). The substitution rates for members of the TG rickettsia and members of the AAE clade for which long spacer sequences are available (*R. akari* and *R. australis*) are comparable (0.20 substitutions per site). We note that *R. canada* is separated from the TG rickettsia by a similar frequency of substitutions per site (0.20). In contrast, *R. australis* and *R. akari* are separated from the SFG rickettsia by two- to threefold lower substitution rates of 0.07 and 0.09 per site, consistent with their phylogenetic placement within the SFG rickettsia. For *R. canada*, the corresponding rate of substitutions is 0.12 per site. The substitution rates for all pairwise comparisons within the classical SFG rickettsia were found to be very small, less than 0.05 substitutions per site.

Repetitive sequences flanking sites of insertions/deletions were identified in the spacer regions of several *Rickettsia* species. These repetitive elements range in size from 7 to 23 bp and are dispersed throughout the intergenic region. For example, a short sequence repeat (TATAGTA) located at the 5′ end of the spacer region in *R. rickettsii* is flanking a 28-bp sequence that is uniquely present in *R. rickettsii*. The three other repeated sequences are located in immediate proximity, with no intervening sequences. For example, the sequence TCATAAGAAATAA is present in two adjacent copies in *R. akari*. The 3′ half of this repeat sequence (ATAAAA) is present in three tandem copies in *R. canada*. A direct repeat with the motif GAAAATAA is located at the 3′ end of the spacer region in the classical SFG rickettsia (*R. conori*, *R. parkeri*, *R. sibirica*, and *R. rickettsii*). Immediately downstream of this sequence repeat, there is a repeat of 23 bp in *R. akari* (TTGTGAAATTAATTCTGTATGCA), but not in any of the other species.

Synonymous and Nonsynonymous Substitutions in the *fmt* Gene

Amino acid sequences of the C-terminal parts of the *fmt* gene products were aligned. The length of the

Table 1
Frequencies of Substitutions (%) Between the 23S rRNA Gene (above diagonal) and the *fmt-rrl* Spacer Sequences (below diagonal)

Species	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1. *Rickettsia conorii*	—	0.29	0.33	0.47	ND	0.76	0.47	1.44	1.63	ND	1.77	1.99	1.99	2.28	10.8
2. *Rickettssia parkeri*	1.05	—	0.40	0.61	ND	0.91	0.61	1.51	1.78	ND	1.73	1.99	1.92	2.28	10.8
3. *Rickettsia sibirica*	1.40	0.82	—	0.58	ND	0.80	0.58	1.40	1.67	ND	1.73	2.03	2.03	2.28	10.9
4. *Rickettsia rickettsii*	2.47	2.10	2.22	—	ND	0.91	0.58	1.48	1.74	ND	1.81	2.10	2.10	2.18	10.9
5. *Rickettsia amblyommii*	3.20	2.84	2.96	3.56	—	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND
6. *Rickettsia montana*	3.43	3.07	2.95	3.79	1.89	—	0.65	1.85	1.89	ND	1.88	2.18	2.17	2.43	11.0
7. *Rickettsia rhipicephali*	3.67	3.31	3.43	3.80	2.13	2.60	—	1.40	1.52	ND	1.66	1.81	1.88	2.14	10.8
8. *Rickettsia australis*	7.57	7.56	7.68	7.83	6.52	6.97	6.74	—	1.33	ND	2.22	2.51	2.62	2.59	11.0
9. *Rickettsia akari*	9.03	8.41	9.02	9.18	8.43	8.66	8.06	7.56	–	ND	2.36	2.40	2.54	2.66	10.9
10. *Rickettsia felis*	6.23	6.56	6.56	6.23	5.92	6.88	5.90	9.18	10.2	—	ND	ND	ND	ND	ND
11. *Rickettsia canada*	12.6	12.3	12.2	12.5	11.4	12.0	12.2	13.4	14.6	9.30	—	2.65	2.50	2.69	10.9
12. *Rickettsia typhi*	20.1	20.1	20.1	20.6	18.5	19.6	19.2	19.0	20.5	14.5	19.9	—	1.12	2.77	10.4
13. *Rickettsia prowazekii*	20.4	20.6	20.6	20.5	19.3	20.1	19.8	19.3	21.0	12.8	21.0	11.0	—	2.77	10.6
14. *Rickettsia bellii*	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	—	10.6
15. *Orientia tstsugamushi*	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	ND	—

NOTE.—ND indicates data not available.

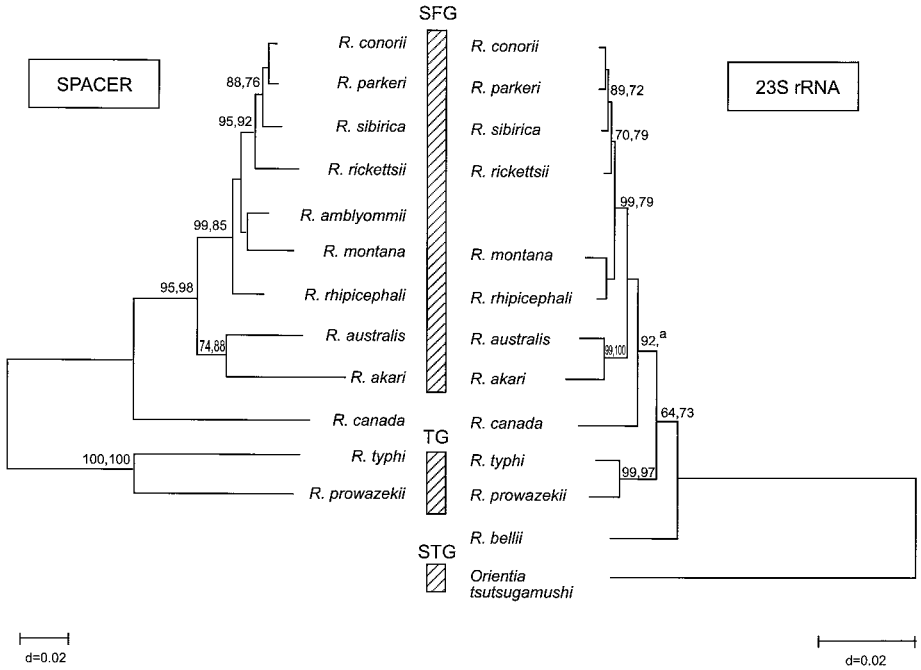

FIG. 3.—Phylogenetic relationships of *Rickettsia* species based on the spacer sequences between *fmt* and *rrl* and based on the 23S gene sequences. Neighbor-joining and maximum-parsimony methods using the phylo_win interface gave similar topologies, as detailed in the text. Branch lengths are proportional to those reconstructed with the neighbor-joining method. Values at nodes indicate the percentages of 500 bootstraps (neighbor joining, parsimony) in which the taxa to the right were clustered. Only bootstrap values above 70% are shown. [a] The branch separating *R. canada* from *R. typhi* and *R. prowazekii* was collapsed in the maximum-parsimony analysis. TG = typhus group; SFG = spotted fever group; STG = scrub typhus group. The sequence alignment is available at the SMBE web site. The url for this site is currently http://hgc.sph.uth.tmc.edu/smbe/.

3′ end of the *fmt* gene was found to be identical in all species, with the exception of *R. felis*, in which three contiguous nucleotides were found to be deleted. Since the deletion corresponds to exactly one codon, no shift in the reading frame is observed. K_S and K_A for the *fmt* gene distinguish the TG rickettsia from the SFG rickettsia in all pairwise comparisons. For these groups, the mean nonsynonymous and the mean synonymous substitution rates are 0.07 and 0.26 per position, respectively. Based on these values, the *fmt* gene of *R. felis* appears to be much more similar to the SFG rickettsia (K_A = 0.01; K_S = 0.06) than to the TG rickettsia (K_A = 0.06; K_S = 0.31). Similarly, *R. canada* is separated from the TG rickettsia by two- to threefold higher substitution rates per site (K_A = 0.10; K_S = 0.23) than for the SFG rickettsia (K_A = 0.06; K_S = 0.08). Within the the TG and SFG rickettsia, synonymous substitution frequencies were found to be less than 0.10 substitutions per position.

Phylogenetic Relationships

In order to trace the evolutionary history of the rearrangement event, we performed phylogenetic reconstructions based on the *fmt-rrl* spacer sequences and the 23S rRNA genes (fig. 3). The evolutionary relationship of species within the genus *Rickettsia* was also examined from a combined data set of more than 3,700 nucleotide sites, including sequences from the *fmt* genes, the *fmt-rrl* spacers, and the 23S rRNA genes (fig. 4). Although there is about a 10-fold difference in sequence divergence between the *fmt-rrl* spacer regions and the *rrl* gene sequences (table 1), the resulting tree topologies are largely congruent and have features as outlined below.

First, *R. prowazekii* and *R. typhi* form a cluster that is distinct from the cluster formed by members of the SFG rickettsia as well as from *R. canada,* which has previously been classified in the TG rickettsia. Within the SFG rickettsia, three main clusters can be identified: *R. australis* and *R. akari* branch together as an early-diverging clade (74%–100% bootstrap support), and *R. conorii, R. parkeri, R. sibirica,* and *R. rickettsii* appear to be phylogenetically distinct from the nonpathogenic species *R. amblyommii, R. rhipicephali,* and *R. montana* (70%–98% bootstrap support).

Second, maximum-parsimony analysis based on the *fmt-rrl* spacer sequences produced seven equally parsimonious trees which differed only in the branching order within the cluster comprising *R. conorii, R. parkeri,* and *R. sibirica* and the cluster containing *R. montana, R. amblyommii,* and *R. rhipicephali.* Whereas there is evidence to suggest that *R. rickettsii* diverged prior to *R. conori, R. parkeri,* and *R. sibirica* (72%–97% bootstrap support), the order of divergence among the latter three could not be resolved in any of the reconstructions. The branching order of *R. montana* and *R. rhipicephali* could not be resolved in the phylogenetic reconstructions based on either the *fmt-rrl* spacer region or the 23S rRNA gene sequences, as inferred from very low

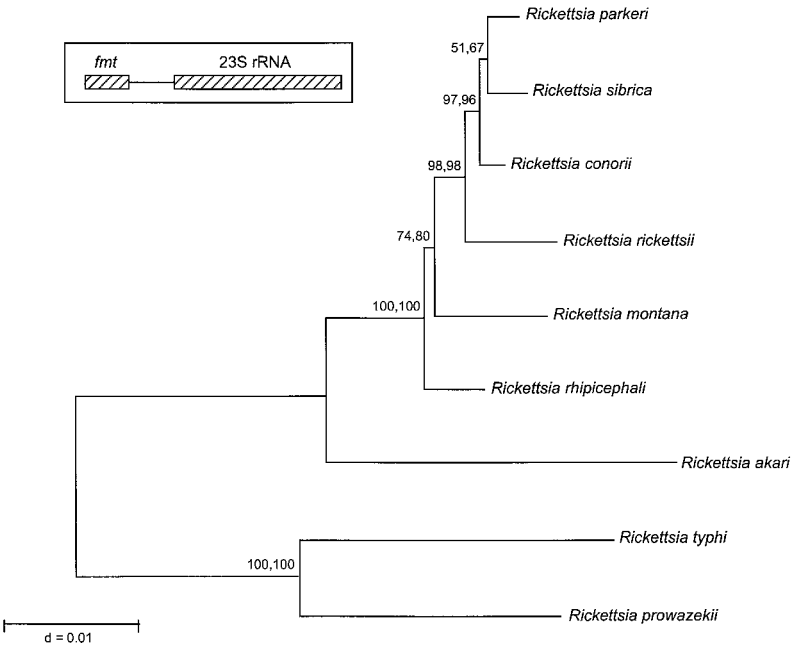

FIG. 4.—Phylogenetic relationships of *Rickettsia* species derived from the combined sequences of the *fmt* genes, the *fmt-rrl* spacer regions, the 23S rRNA genes and ca. 100 bases of the *rrl-rrf* spacer regions (3,735 homologous sites). Neighbor-joining and maximum-parsimony methods using the phylo_win interface gave identical topologies. Branch lengths are proportional to those reconstructed with the neighbor-joining method. Values at nodes indicate the percentages of 500 bootstraps (neighbor joining, parsimony) in which the taxa to the right were clustered.

bootstrap score values. However, the phylogenetic analysis based on the combined data set provides some evidence to suggest that *R. rhipicephali* diverged prior to *R. montana* (74% and 80% bootstrap support for neighbor joining and maximum parsimony, respectively).

Finally, the neighbor-joining tree based on the 23S rRNA gene sequences indicates that *R. canada* diverged from the SFG after the SFG-TG schism (92% bootstrap support); however, this internal branch collapsed in the maximum-parsimony analysis. *Rickettsia felis* was not included in the analysis because of the 500 bp deletion. Phylogenetic trees based on the remaining 280 aligned sites in the *fmt-rrl* spacer regions are less well supported than the tree presented in figure 3, but they indicate that *R. felis* may have diverged prior to *R. australis* and *R. akari*, but subsequent to *R. canada* (data not shown). Taken together, our phylogenetic analyses based on the *fmt-rrl* spacer and on the 23S rRNA gene sequences are largely congruent with trees generated from the 16S rRNA, the 17-kDa and the citrate synthase gene sequences (Anderson et al. 1987; Stothard and Fuerst 1995; Roux and Raoult 1995*b*; Roux et al. 1997), with some ambiguity as to the placement of *R. canada.*

Discussion

Rearranged rRNA operon structures are rare in bacterial genomes. So far, all exceptions to the standard rRNA operon structure (16S-23S-5S) (fig. 1*A*) are associated with small genomes that have low copy numbers of the rRNA genes (Andersson and Kurland 1995). A few such exceptions have been observed within the mycoplasmas, where genome sizes range from 580 kb to 1 Mb. For example, in *Mycoplasma hyopneumoniae*, the single 5S rRNA gene is separated from the single 16S and the 23S rRNA genes (Taschke and Herrmann 1986). In *Mycoplasma gallisepticum*, there are two copies of a 23S-5S rRNA gene cluster, none of which is located near the single 16S rRNA gene (Chen and Finch 1989). Surprisingly, the smallest genome within this group, *Mycoplasma genitalium*, has a single set of the rRNA genes that are arranged in the standard manner (Fraser et al. 1995). This variability in gene order structures implies a high degree of genomic plasticity in the small *Mycoplasma* genomes.

The 1.1 Mb *R. prowazekii* genome has recently been sequenced (Andersson et al. 1998). This genome is thought to have evolved from a much larger genome of a free-living ancestor that learned how to exploit the eukaryotic cell as its growth environment (Weisburg et al. 1989; Andersson and Kurland 1995). The resulting reduction in genome size is believed to have been associated with extensive genomic rearrangement events, possibly caused by intrachromosomal recombination at repeated sequences and duplicated genes (Andersson and Kurland 1995). Indeed, the genomic neighborhoods of the genes encoding rRNAs and elongation factors suggest that numerous rearrangement events have occurred during the evolution of the *R. prowazekii* genome (Andersson et al. 1995; Syvänen et al. 1996).

In a first attempt to backtrack through the evolutionary history of *R. prowazekii* to identify the points at which the rearrangement of the rRNA genes occurred, we characterized the organization of the 23S rRNA genes in 14 closely related species of *Rickettsia.* In 12

of these, we found a gene organization pattern that is identical to that observed in *R. prowazekii,* i.e., the gene encoding *fmt* is located immediately upstream of the gene coding for 23S rRNA (fig. 2). However, we were unable to identify the *fmt* gene upstream of the 23S rRNA gene in *R. bellii* and *O. tsutsugamushi* using the PCR techniques described here. This suggests either that the *fmt* genes in these species have diverged too much to be recognized by the primers used or that the *fmt* gene is not located upstream of the 23S rRNA gene in these species. Previous attempts to show linkage between the 16S and the 23S rRNA genes in *R. bellii* using PCR and Southern analysis were unsuccessful (Clark 1990; Stothard 1995). Therefore, it seems likely that *R. bellii* also has a disrupted operon, but it remains to be determined whether the structure is the same as that in the other species. In contrast to the rearrangement of the 16S and 23S rRNA genes, the 23S and 5S rRNA genes are linked in all of the *Rickettsia* species examined, as inferred from the results of PCR reactions and partial sequence analyses (data not shown).

We used the 23S rRNA gene and the *fmt-rrl* spacer region to reconstruct the phylogeny of *Rickettsia* in order to place the rearrangement event of the 23S rRNA genes in an evolutionary context. The phylogenetic analysis provides strong support for many of the clusters previously suggested based on 16S rRNA and citrate synthase gene sequences (Stothard and Fuerst 1995; Roux et al. 1997). For example, members of the TG rickettsia (*R. prowazekii* and *R. typhi*) are clearly separated from members of the SFG rickettsia. The analyses also strongly support the suggestion that *R. akari* and *R. australis* represent an early-diverging group within the SFG rickettsia and that *R. bellii* diverged prior to the split between the TG and SFG rickettsia (Stothard, Clark, and Fuerst 1994; Stothard and Fuerst 1995; Roux et al. 1997). In contrast to these well-supported groups, there are some discrepancies regarding the placement of *R. canada* in the various gene trees that have been produced. According to the neighbor-joining 23S rRNA trees presented here, this species may represent the earliest-diverging species within the SFG rickettsia. However, phylogenetic reconstructions based on the *gltA* gene suggest that it diverged prior to the split between the TG and the SFG rickettsia but subsequent to the divergence of *R. bellii* (Roux et al. 1997). Nevertheless, the trees based on the 23S rRNA gene sequences and the *fmt-rrl* spacer sequences are largely congruent with previously suggested phylogenetic relationships among the *Rickettsia.*

By superimposing the observed rearrangement of the rRNA genes onto the derived consensus tree (fig. 4), we conclude that the rearrangement of the rRNA genes preceded the divergence of the TG and the SFG rickettsia. However, since a complete rRNA operon cannot be found in *R. bellii,* we cannot say with certainty whether the rearrangement occurred at the time of or prior to the speciation event leading to the genus *Rickettsia.* If *R. canada* diverged prior to the SFG-TG schism, as the citrate synthase gene trees suggest, then there is evidence that the rearrangement event did in fact occur early in the history of the genus. However, if *R. canada* diverged from within the SFG lineage, as the 23S rRNA neighbor-joining phylogeny suggests, it is more difficult to establish how early the rearrangement event occurred. Other α-proteobacterial species, such as, for example, *Bartonella* sp., have rRNA operons that are typical of bacteria, i.e., the 16S rRNA gene is located immediately upstream of the 23S rRNA gene (Roux and Raoult 1995*a*). Examination of rRNA operon structures in other Rickettsiales genera is clearly necessary to identify the exact node at which the rearrangement event occurred.

Mutational degradation and differential deletion of individual 16S and 23S rRNA genes within multiple rRNA gene operons is one mechanism whereby unlinked, single-copy rRNA genes could have been generated. Indeed, the *R. prowazekii* genome contains an unusually high fraction of noncoding DNA (ca. 24%), part of which may be remnants of old genes that are in the process of being discarded from the genome (Andersson et al. 1998). A detailed comparative study of pseudogenes from a variety of *Rickettsia* species indicates that elimination of gene sequences is a gradual and ongoing process in *Rickettsia* (Andersson and Andersson 1999). However, since the spacer regions upstream of the 23S rRNA gene show no sequence similarities to genes in the public databases, we have no clues to suggest what the *fmt-rrl* spacer region might once have been coding for, if anything.

The rearrangement of the rRNA genes may have been mediated by recombination events between directly repeated sequences that resulted in deletions of the intervening segments. Here, we note that the upstream spacer regions of the rearranged 23S rRNA genes in some *Rickettsia* species contain short repetitive sequences that have been eliminated in other related species. The scattered distribution of adjacent repeat sequences in the spacer regions suggests that they may have been eliminated by intrachromosomal recombination events several times independently during the evolution of the *Rickettsia* lineages. Further comparative genomic data will be needed to quantitate the relative frequencies of rearrangement and deletion events within the rickettsial genomes.

Acknowledgments

We thank Magali Romedenne, Natacha Viseur, and Cecilia Alsmark for technical assistance. This work was supported by the Department of Molecular Genetics at the Ohio State University (P.F.), the Natural Sciences Research Council, and the Knut and Alice Wallenberg Foundation (S.G.E.A. and C.G.K.).

LITERATURE CITED

ANACKER, R. L., R. E. MANN, and C. GONZALES. 1987. Reactivity of monoclonal antibodies to *Rickettsia rickettsii* with spotted fever and typhus group rickettsiae. J. Clin. Microbiol. **25**:167–171.

ANDERSON, B. E., R. L. REGNERY, G. M. CARLONE, T. TZIANABOS, J. E. MCDADE, Z. Y. FU, and W. J. BELLINI. 1987. Sequence analysis of the 17-kilodalton-antigen gene from *Rickettsia rickettsii.* J. Bacteriol. **169**:2385–2390.

ANDERSSON, J. O., and S. G. E. ANDERSSON. 1999. Genome degradation is an ongoing process in *Rickettsia.* Mol. Biol. Evol. (in press).

ANDERSSON, S. G. E., and C. G. KURLAND. 1995. Genomic evolution drives the evolution of the translation system. Biochem. Cell Biol. **73**:775–787.

ANDERSSON, S. G. E., and P. M. SHARP. 1996. Codon usage and base composition in *Rickettsia prowazekii.* J. Mol. Evol. **42**:525–536.

ANDERSSON, S. G. E., A. ZOMORODIPOUR, J. O. ANDERSSON, T. SICHERITZ-PONTÉN, U. C. ALSMARK, R. F. PODOWSKI, A. K. NÄSLUND, A.-S. ERIKSSON, H. H. WINKLER, and C. G. KURLAND. 1998. The genome sequence of *Rickettsia prowazekii* and the origin of mitochondria. Nature **396**:133–140.

ANDERSSON, S. G. E., A. ZOMORODIPOUR, H. H. WINKLER, and C. G. KURLAND. 1995. Unusual organization of the rRNA genes in *Rickettsia prowazekii.* J. Bacteriol. **177**:4171–4175.

AZAD, A. F., J. B. SACCI JR., W. M. NELSON, G. A. DASCH, E. T. SCHMIDTMAN, and M. CARL. 1992. Genetic characterization and transovarial transmission of a typhus like rickettsia found in cat fleas. Proc. Natl. Acad. Sci. USA **89**:43–46.

BLATTNER, F. R., G. PLUNKETT III, C. A. BLOCH et al. (16 co-authors). 1997. The complete genome sequence of *Escherichia coli* K12. Science **277**:1453–1462.

BURGDORFER, W., and L. P. BRINTON. 1970. Intranuclear growth of *Rickettsia canada,* a member of the typhus group. Infect. Immun. **30**:112–114.

CABOT, E. L., and A. T. BECKENBACH. 1989. Simultaneous editing of multiple nucleic acid and protein sequences with ESEE. Comput. Appl. Biosci. **5**:233–234.

CHEN, X., and L. R. FINCH. 1989. Novel arrangement of rRNA genes in *Mycoplasma gallisepticum*: separation of the 16S gene of one set from the 23S and 5S genes. J. Bacteriol. **171**:2876–2878.

CLARK, J. 1990. A molecular phylogenetic analysis of intracellular bacteria Rickettsiae. Dissertation, Ohio State University, Columbus.

EREMEEVA, M. E., V. ROUX, and D. RAOULT. 1993. Determination of genome size and restriction pattern polymorphism of *Rickettsia prowazekii* and *Rickettsia typhi* by pulsed field gel electrophoresis. FEMS Microbiol. Lett. **112**:105–112.

FRASER, C. M., J. D. GOCAYNE, O. WHITE et al. (29 co-authors). 1995. The minimal gene complement of *Mycoplasma genitalium.* Science **270**:397–403.

GALTIER, N., M. GOUY, and C. GAUTIER. 1996. SeaView and Phylo_win, two graphic tools for sequence alignment and molecular phylogeny. Comput. Appl. Biosci. **12**:543–548.

GUTELL, R. R., M. W. GRAY, and M. N. SCHNARE. 1993. A compilation of large subunit (23S and 23S-like) ribosomal RNA structures: 1993. Nucleic Acids Res. **21**:3055–3074.

KRAWIEC, S., and M. RILEY. 1990. Organization of the bacterial chromosome. Microbiol. Rev. **54**:502–539.

LI, W.-H. 1993. Unbiased estimation of the rates of synonymous and nonsynonymous substitution. J. Mol. Evol. **36**: 96–99.

MCKEIL, J. A., E. J. BELL, and D. B. LACKMAN. 1967. *Rickettsia canada*: a new member of the typhus group rickettsiae isolated from *Leporis palustris* ticks in Canada. Can. J. Microbiol. **13**:503–510.

OLSEN, G. J., C. R. WOESE, and R. OVERBEEK. 1994. The winds of (evolutionary) change: breathing new life into microbiology. J. Bacteriol. **175**:3893–3896.

PHILIP, R. N., E. A. CASPER, R. L. ANACKER, J. CORY, S. F. HAYES, W. BURGDORFER, and C. E. YUNKER. 1983. *Rickettsia bellii* sp. no.: a tick borne *Rickettsia,* widely distributed in the United States, that is distinct from the spotted fever and typhus biogroups. Int. J. Syst. Bacteriol. **33**:94–106.

PRETZMAN, C. I., Y. RIKIHISA, D. RALPH, J. C. GORDON, and S. BECH-NIELSEN. 1987. Enzyme-linked immunosorbent assay for Potomac horse fever disease. J. Clin. Microbiol. **25**: 31–36.

ROUX, V., and D. RAOULT. 1995*a.* The 16S-23S rRNA intergenic spacer region of *Bartonella* (*Rochalimaea*) species is longer than usually described in other bacteria. Gene **156**: 107–111.

———. 1995*b.* Phylogenetic analysis of the genus *Rickettsia* by 16S rDNA sequencing. Res. Microbiol. **146**:385–396.

ROUX, V., E. RYDKINA, M. EREMEEVA, and D. RAOULT. 1997. Citrate synthase gene comparison, a new tool for phylogenetic analysis, and its application for the rickettsiae. Int. J. Syst. Bacteriol. **47**:252–261.

SAITOU, N., and M. NEI. 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. **4**:406–425.

STOTHARD, D. R. 1995. The evolutionary history of the genus *Rickettsia* as inferred from 16S and 23S rRNA genes and the 17 kDa cell surface antigen gene. Dissertation, Ohio State University, Columbus.

STOTHARD, D. R., J. B. CLARK, and P. A. FUERST. 1994. The phylogenetic relationship of *Rickettsia bellii* to the spotted fever and typhus group of *Rickettsia,* and the antiquity of the genus. Int. J. Syst. Bacteriol. **44**:798–804.

STOTHARD, D. R., and P. A. FUERST. 1995. Evolutionary analysis of the spotted fever and typhus groups of *Rickettsia* using 16S rRNA gene sequences. Syst. Appl. Bacteriol. **18**: 52–61.

SYVÄNEN, A.-C., H. AMIRI, A. JAMAL, S. G. E. ANDERSSON, and C. G. KURLAND. 1996. A chimeric disposition of the elongation factor genes in *Rickettsia prowazekii.* J. Bacteriol. **178**:6192–6199.

TAMURA, A., N. OHASHI, H. URAKAMI, and 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.

TASCHKE, C., and R. HERRMANN. 1986. Analysis of transcription and processing signals of the 16S-23S rRNA operon of *Mycoplasma hyopneumoniae.* Mol. Gen. Genet. **205**: 4705–4707.

THOMPSON, J. D., D. G. HIGGINS, and T. J. GIBSON. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res. **22**:4683–4680.

WEISBURG, W. G., M. E. DOBSON, J. E. SAMUEL, G. A. DASCH, L. P. MALLAVIA, O. BACA, L. MANDELCO, J. E. SECHRESTR, E. WEISS, and C. R. WOESE. 1989. Phylogenetic diversity of the Rickettsiae. J. Bacteriol. **171**:4202–4206.

WEISS, E., and J. M. MOULDER. 1984. Order I. Rickettsiales. Pp. 687–704 *in* N. R. KREIG and J. G. HOLT, eds. Bergey's manual of systematic bacteriology. Williams and Wilkins, Baltimore.

WILLIAMS, S. G., J. B. SACCI JR., M. E. SCHRIEFER, E. M. ANDERSEN, K. K. FUJIOKA, F. J. SORVILLO, A. R. BARR, and A. F. AZAD. 1992. Typhus and typhus-like rickettsiae associated with opossums and their fleas in Los Angeles County, California. J. Clin. Microbiol. **30**:1758–1762.

WISSEMAN, C. L. JR., E. A. EDLINGER, A. D. WADDELL, and M. R. JONES. 1976. Infection cycle of *Rickettsia rickettsii* in chick embryo and L-929 cells in culture. Infect. Immun. **14**:1052–1064.

DAN GRAUR, reviewing editor

Accepted March 22, 1999