Evolutionary Relationships among Members of the Genus Chlamydia Based on 16S Ribosomal DNA Analysis

BERTIL PETTERSSON,1 ANNALENA ANDERSSON,1 THOMAS LEITNER,2 ÖRJAN OLSVIK,3 MATHIAS UHLEN,1 CHRISTOPHER STOREY,4 AND CAROLYN M. BLACK5*

Department of Biochemistry and Biotechnology, The Royal Institute of Technology, Stockholm, Sweden;1 Theoretical Biology and Biophysics, Group T-10, Los Alamos National Laboratory, Los Alamos, New Mexico;2 Department of Medical Microbiology, School of Medicine, University of Tromsø, Tromsø, Norway;3 Department of Microbiology, University of Leeds, Leeds, United Kingdom;4 and National Center for Infectious Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia5

Received 31 October 1996/Accepted 1 May 1997

Nucleotide sequences from strains of the four species currently in the genus Chlamydia, C. pneumoniae, C. psittaci, and C. trachomatis were investigated. In vitro-amplified RNA genes of the ribosomal small subunit from 30 strains of C. pneumoniae and C. pecorum were subjected to solid-phase DNA sequencing of both strands. The human isolates of C. pneumoniae differed in only one position in the 16S rRNA gene, indicating genetic homogeneity among these strains. Interestingly, horse isolate N16 of C. pneumoniae was found to be closely related to the human isolates of this species, with a 98.9% nucleotide similarity between their 16S rRNA sequences. The type strain and koala isolates of C. pecorum were also found to be very similar to each other, possessing two different 16S rRNA sequences with only one-nucleotide difference. Furthermore, the C. pecorum strains truncated the 16S rRNA molecule by one nucleotide compared to the molecules of the other chlamydial species. This truncation was found to result in loss of a unilaterally bulged nucleotide, an attribute present in all other eubacteria. The phylogenetic structure of the genus Chlamydia was determined by analysis of 16S rRNA sequences. All phylogenetic trees revealed a distinct line of descent of the family Chlamydiaceae built of two main clusters which we denote the C. pneumoniae cluster and the C. psittaci cluster. The clusters were verified by bootstrap analysis of the trees and signature nucleotide analysis. The former cluster contained the human isolates of C. pneumoniae and equine strain N16. The latter cluster consisted of C. psittaci, C. pecorum, and C. trachomatis. The members of the C. pneumoniae cluster showed tight clustering and strain N16 is likely to be a subspecies of C. pneumoniae since these strains also share some antigenic cross-reactivity and clustering of major outer membrane protein gene sequences. C. psittaci and strain N16 branched early out of the respective cluster, and interestingly, their inclusion bodies do not stain with iodine. Furthermore, they also share less reliable features like normal elementary body morphology and plasmid content. Therefore, the branching order presented here is very likely a true reflection of evolution, with strain N16 of the species C. pneumoniae and C. psittaci forming early branches of their respective cluster and with C. trachomatis being the more recently evolved species within the genus Chlamydia.
genetic and evolutionary studies of bacteria (31, 44). The phylogenetic relationship of \( C. \) pneumoniae to \( C. \) trachomatis and \( C. \) psittaci has been reported by Gaydos et al. (13). Based on analysis of 16S ribosomal DNA (rDNA) sequences of two strains of \( C. \) pneumoniae, Gaydos et al. found that \( C. \) pneumoniae was more closely related to \( C. \) psittaci than to \( C. \) trachomatis. In a preliminary comparison performed by our laboratory of strains of \( C. \) pneumoniae of diverse geographic origins, three types of 16S rDNA genes were identified (5).

The DNA sequence of the 16S rRNA gene and the characterization of the primary structure of the 16S rRNA molecule have not previously been published for the most recently described species, \( C. \) pecorum. This species consists of a group of ruminant strains associated with infectious polyarthritis, encephalitis, pneumonia, and diarrhea (10, 11). Studies reporting that DNA-DNA similarity between \( C. \) pecorum and strains of other \( Chlamydia \) species is less than 10% (10), immunological specificity of the major outer membrane protein (MOMP) gene (10), and dendograms based on multistain comparison of the MOMP gene sequence (25) have supported the classification of this new species \( C. \) pecorum.

In this report, we present sequence data representing approximately 90% of the 16S rRNA genes from 26 \( C. \) pneumoniae strains and 4 \( C. \) pecorum strains obtained by semiautomated solid-phase DNA sequencing (20, 21, 32–34, 41), and we provide new information on phylogenetic relationships among all of the \( Chlamydia \) species. Signature nucleotides which indicate the stability of the phylogeny and provide a guide for development of PCR-based diagnostic systems based on the 16S rRNA genes of these important bacteria are also presented. The phylogeny of the genus \( Chlamydia \) was inferred from 16S rRNA sequences determined in this work and from previously deposited sequences for strains of \( C. \) trachomatis and \( C. \) psittaci.

### MATERIALS AND METHODS

**Bacterial strains and growth conditions.** The \( Chlamydia \) strains used in this study are listed in Table 1. All strains of \( C. \) pneumoniae were cultivated in HEp-2 cells, harvested, and purified as described previously (1, 3). All DNA templates made from \( C. \) pneumoniae strains for PCR amplification were prepared as described previously (3), using alkaline lysis and proteinase K digestion. \( C. \) pneumoniae strains were cultivated in McCoy cells and harvested as described previously (1, 3). The primers 589 and 590 were used for PCR amplification of the 16S rRNA genes (3). The PCRs were performed directly from genomic material with 5 pmol of each primer. The thermocycling profile

![Table 1. Strains of Chlamydia included in this study](image-url)

<table>
<thead>
<tr>
<th>Species</th>
<th>Strain</th>
<th>Location of origin</th>
<th>Source</th>
<th>Acc. no.</th>
</tr>
</thead>
<tbody>
<tr>
<td>( C. ) pneumoniae</td>
<td>TW-183</td>
<td>Taiwan</td>
<td>Washington Research Foundation</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>CWL-011</td>
<td>Atlanta, Ga.</td>
<td>CDC</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>CWL-029 (VR1310)</td>
<td>Atlanta, Ga.</td>
<td>CDC (ATCC)</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>CWL-050</td>
<td>Atlanta, Ga.</td>
<td>CDC</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>CM-1 (VR1360)</td>
<td>Atlanta, Ga.</td>
<td>CDC (ATCC)</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>IOL-207</td>
<td>Taiwan</td>
<td>P. Nicolini</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>AR-388</td>
<td>Seattle, Wash.</td>
<td>T. Grayston</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>AR-39</td>
<td>Seattle, Wash.</td>
<td>T. Grayston</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>T2023 (VR1356)</td>
<td>Brooklyn, N.Y.</td>
<td>M. Hammerschlag (ATCC)</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>T2043 (VR1355)</td>
<td>Brooklyn, N.Y.</td>
<td>M. Hammerschlag (ATCC)</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>FML-07</td>
<td>Norway</td>
<td>B. P. Berdal</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>FML-12</td>
<td>Norway</td>
<td>B. P. Berdal</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>FML-16</td>
<td>Norway</td>
<td>B. P. Berdal</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>H-12</td>
<td>Finland</td>
<td>P. Saikku</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>1 (Parola)</td>
<td>Finland</td>
<td>P. Saikku</td>
<td>U73783</td>
</tr>
<tr>
<td></td>
<td>K6</td>
<td>Finland</td>
<td>P. Saikku</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>K66</td>
<td>Finland</td>
<td>P. Saikku</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>UZG1</td>
<td>Belgium</td>
<td>J. Ossewaarde</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>WI-02</td>
<td>Madison, Wis.</td>
<td>B. MacDonald</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>WI-03</td>
<td>Madison, Wis.</td>
<td>B. MacDonald</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>WI-04</td>
<td>Madison, Wis.</td>
<td>B. MacDonald</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>WI-05</td>
<td>Madison, Wis.</td>
<td>B. MacDonald</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>WI-06</td>
<td>Madison, Wis.</td>
<td>B. MacDonald</td>
<td>id. to U73783</td>
</tr>
<tr>
<td></td>
<td>YK-41</td>
<td>Hiroshima, Japan</td>
<td>Y. Kanamoto</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>TWIS</td>
<td>Madison, Wis.</td>
<td>R. Adudge</td>
<td>id. to L06108</td>
</tr>
<tr>
<td></td>
<td>N16</td>
<td>Manchester, United Kingdom</td>
<td>C. Storey</td>
<td>U73784</td>
</tr>
</tbody>
</table>

\( a \) All strains are human isolates unless indicated otherwise. \( C. \) trachomatis L2/434 (acc. no. M59178) and SFPD (acc. no. M83313), \( C. \) psittaci 6BC (acc. no. M13769) and OEA (acc. no. Z49871), and a Simkania strain (acc. no. L27666) were included for comparison in the phylogenetic analyses based on sequences available from GenBank but were not sequenced in this study.

\( b \) CDC, Centers for Disease Control and Prevention.

\( c \) ATCC, American Type Culture Collection.

\( d \) id. to, identical to.

\( e \) ATCC, American Type Culture Collection.

\( f \) Bovine isolate.

\( g \) Koala isolate.
TABLE 2. Primers used for PCR and DNA sequencing of the chlamydial strains studied in this work

<table>
<thead>
<tr>
<th>Primer</th>
<th>Position</th>
<th>Sequence</th>
<th>Application</th>
</tr>
</thead>
<tbody>
<tr>
<td>S89</td>
<td>75–94</td>
<td>5′-ATAATGACTCGGTGTTGAT-3′</td>
<td>PCR</td>
</tr>
<tr>
<td>S90B</td>
<td>1455–1458</td>
<td>5′-TATAAAATGTTGTAGC-3′</td>
<td>PCR</td>
</tr>
<tr>
<td>S89BE</td>
<td>75–94</td>
<td>5′-ATAATGACTCGGTGTTGAT-3′</td>
<td>Sequencing</td>
</tr>
<tr>
<td>S58F</td>
<td>800–818</td>
<td>5′-TTAGTCACGCCGTAACG-3′</td>
<td>Sequencing</td>
</tr>
<tr>
<td>S90F</td>
<td>939–947</td>
<td>5′-CTTGGCGGYYCCGTCAATTC-3′</td>
<td>Sequencing</td>
</tr>
<tr>
<td>S57F</td>
<td>1175–1193</td>
<td>5′-GAGGAAGGTGGGGATGATG-3′</td>
<td>Sequencing</td>
</tr>
<tr>
<td>S590F</td>
<td>1455–1458</td>
<td>5′-TATAAAATGTTGTAGC-3′</td>
<td>Sequencing</td>
</tr>
</tbody>
</table>

* a The letters B and F indicate that a biotin dye (B) or a fluorescein isothiocyanate (F) dye is coupled to the oligonucleotide.

* b Numbers are positions according to the International Union of Biochemistry numbering of the 16S rRNA gene of E. coli (6).

* c Y denotes degenerate positions according to the International Union of Biochemistry letter code.

RESULTS AND DISCUSSION

Nucleotide sequences of 16S rRNA. Direct solid-phase DNA sequencing resulted in an unambiguous determination of about 90% of the 16S rRNA genes in both directions from the Chlamydia strains listed in Table 1. Because internal primers were used to amplify the target gene, the sequences are incomplete at their 5′ and 3′ ends. The sequences obtained ranged between 1,354 nucleotides (nt) for the C. psittaci strains to 1,355 nt for the C. pneumoniae strains. Microheterogeneities (polymorphisms) due to nucleotide differences between the rRNA genes at certain positions were not observed, as recently found in certain strains of mycoplasmas and Bacillus sporothermodurans (32–34). This indicated that only one 16S rRNA gene is present in C. pneumoniae and C. pecorum, which is in agreement with recent findings (12). Moreover, the 16S rRNA sequences obtained from the strains of C. pneumoniae and C. pecorum showed a very high intraspecies similarity. The human isolates of C. pneumoniae showed the presence of two different nucleotide sequences. The difference was due to a variation in only one nucleotide position. Seventeen of the C. pneumoniae strains were found to have a G in position 490 according to the Escherichia coli numbering (6), and this sequence variant is represented by C. pneumoniae P1 (Parola). The sequences of the remaining eight C. pneumoniae strains of this species originating from humans were identical to the previously published one for this species (i.e., C. pneumoniae IOL207) and therefore not deposited. Interestingly, the equine isolate of C. pneumoniae, strain N16, differed in 13 and 14 positions, respectively, compared with the human strains of this species. At position 1324, the type strain Bo/E8 of C. pecorum had a G, while the other three strains of this species had an adenosine residue. The latter variant is represented by C. pecorum BE.

Additional chlamydial 16S rRNA sequences from strains other than those described in Materials and Methods were deposited in GenBank during the preparation of this manuscript. These sequences were subsequently retrieved for inclusion in our phylogenetic analysis. Since they were found to differ significantly in length, they were not included in the data set for subsequent computation of the final evolutionary distance matrix (see Table 3 and phylogenetic tree (Fig. 1), with the exception of the near-full-length 16S rRNA gene sequence from “Chlamydia” sp. strain bn9 deposited into GenBank under the accession number Y07556.

Despite these length variations, trees were constructed from sets of short sequences (570 and 850 nt) (data not shown) for comparison with the final tree constructed from the full-length sequences determined in this study and shown in Fig. 1. Results of these comparisons did not alter the cluster formations presented in Fig. 1. The strains recently deposited by other investigators with accession numbers (acc. no.) are as follows: C. psittaci GPIC, acc. no. U61768 (570 nt); C. psittaci strain from a pigeon, acc. no. U61767 (570 nt); C. psittaci EAE, acc. no. Z49871 (1,466 nt); C. psittaci W73, acc. no. U61769; C. pecorum VR628, acc. no. U61770 (570 nt); unknown strain, acc. no. I12376 (851 nt); unknown strain, acc. no. I12375 (851 nt); unknown strain, acc. no. I12374 (851 nt); and finally, a strain of C. pneumoniae isolated from a coronary atheroma lesion, acc. no. S56213 (851 nt), published by Kuo et al. (29). Moreover, these 16S rRNA sequences were aligned and used for signature nucleotide and unique nucleotide analysis. Their phylogenetic positions are discussed below according to the cluster to which they were found to belong. To facilitate the discussion of characteristic nucleotide positions and cluster analysis, the chlamydial 16S rRNA gene sequences of the species which are represented in Fig. 1 have been supplied as an alignment to E. coli in Fig. 2.

Phylogenetic analysis of the genus Chlamydia. An uncorrected similarity matrix based on the 16S nucleotide sequences obtained in this work and deposited data from Chlamydia strains, Flexistipes sinusarabici, and selected members of the Planctomyces and relatives is shown in the lower triangle of Table 3. The upper triangle of Table 3 contains distance values corrected for nucleotide events by the method of Jin and Nei (23). Since a 16S rRNA sequence from a strain of the genus Simkania and a strain designated “Chlamydia” sp. strain bn9 (acc. no. Y07556) could be retrieved from GenBank, they were included in the phylogenetic analysis. However, they will not be discussed in detail in this work. Overall, the strains belonging
to the Chlamydia genus were found to be rather similar to each other by 95% or more except for strain bn9 (Table 3). Furthermore, the data presented in Table 3 indicate that the Chlamydia spp. are only distantly related to the genus Simkania due to similarity values less than 85.5%. Investigation of the data set revealed that the nucleotide frequencies had a fairly even distribution (average values were 25.7% for A, 21.9% for C, 30.7% for G, and 21.7% for T) and that the average ts/tv rate bias was approximately 1.0. When no bias occurs, the uncorrected value is 0.5. Furthermore, substantial rate variation across sites was observed and described by a gamma distribution (46). The Jin and Nei substitution model was employed to calculate pairwise distances, since it is the simplest model that can include all these parameters. However, although the distance estimates change with the substitution model used, the topology found with the neighbor-joining method did not change.

Figure 1 shows the topology found by maximum-likelihood and maximum-parsimony calculations. The branch lengths were derived by maximum-likelihood calculations with the Jin and Nei substitution model, and the presented bootstrap values were calculated by maximum parsimony from a set of 500 resamplings. The Chlamydia species formed a distinct line of descent with an early branch of the Simkania strain and the strain designated “Chlamydia” sp. strain bn9. The latter two strains constituted a common cluster in all of the constructed trees (Fig. 1). As judged from the 16S rRNA similarity value (>90%) in Table 3, they are most likely to represent strains belonging to different genera. In addition, since “Chlamydia” sp. strain bn9 (acc. no. Y07556) was <88.4% similar to the four classified species of Chlamydia, the status of this strain in the genus remains to be confirmed. All trees, regardless of substitution model and tree-building method, displayed two main clusters. These clusters are denoted the C. pneumoniae cluster and the C. psittaci cluster and will be discussed in more detail below. The bootstrap analysis indicated a lower value (53%) for one of the forks within the C. psittaci cluster (Fig. 1). Moreover, the branching order within the C. psittaci cluster differed when analyzed by the neighbor-joining method compared to the tree derived by the maximum-likelihood and maximum-parsimony methods (Fig. 1). This was because the subcluster of C. pecorum grouped together with the C. psittaci subcluster instead of that of C. trachomatis. Not surprisingly, the actual branching point within the C. psittaci cluster in the neighbor-joining tree was supported by a low bootstrap value (40%). Thus, the position of the C. pecorum strains altered depending on the tree construction method used for phylogenetic computation. The neighbor-joining method tries to find the minimum-evolution tree (37), and in this analysis, the branching order of the resulting topology using the neighbor-joining method was found to have a slightly better score (i.e., shorter total tree length) than the tree topology found by maximum-parsimony and maximum-likelihood analyses. On the other hand, the neighbor-joining tree was found to be two steps longer in parsimony analysis (1,450 versus 1,448 steps) and involved a little more homoplasy (less consistency among nucleotide sites). Furthermore, by comparing the two alternative branching orders by a log likelihood ratio test under the Jin and Nei model using maximum-likelihood calculations, the topology suggested by the neighbor-joining method was rejected at a P = 0.05 level. To further investigate this difference, we decided to analyze the individual 16S rRNA sequences for unique nucleotide positions in order to confirm the branching order. The reliability of the branching order within the C. psittaci cluster will be further discussed below.
### Signature nucleotides and secondary structure analysis of the 16S rRNA molecule

Signature nucleotides in the 16S rRNA gene of the *Chlamydia* species are presented in Table 4. The nucleotide positions are numbered according to the corresponding positions in the 16S rRNA sequence of *E. coli* (6). Shown in Table 4 is the residue found in all *Chlamydia* strains, the base(s) or base pair(s) found in other taxa, and a list with exceptions (i.e., taxa which share the *Chlamydia* signature). The large number of signatures facilitates development of DNA probes which are *Chlamydia* genus specific. Furthermore, a few positions were found to have no exception, thereby following the strict definition of a signature nucleotide. This is rarely seen when performing signature nucleotide analysis on other taxa. One striking motif was found to be a transition occurring at position 526 where the *Chlamydia* species, including “*Chlamydia*” sp. strain bn9 (acc. no. Y07556) and the *Simkania* strain, all have a uridine residue instead of a cytidine which is present in almost all other organisms. The uridine is situated in the 5'-end region of the 16S rRNA molecule. The signature nucleotide at position 526 was found to be the G:C pairs at positions 290 and 310 which were present in all chlamydiae but only in *Chlamydia psittaci* and *Chlamydia pneumoniae* TW-183 (30, 36). This comparison, performed in 1986, was based on analysis of only 400 sequences of 16S rRNA molecules. Since that time, the total amount of deposited 16S rRNA gene sequences available for retrieval has increased almost 10-fold. However, we found that a similar number of signature positions still characterize the chlamydiae and planctomycetes. The branch of the planctomyces and relatives was found to be supported by 12 signature nucleotides in our analysis. Most of these signatures were analyzed and discussed thoroughly by Weisburg and coworkers due to their effect on the secondary structure of 16S rRNA and occurrence in other taxa (42). The one exception was found to be the G:C pairs at positions 290 and 310 which were present in all chlamydiae but only in *P. staleyi* among the planctomyces and relatives. Interestingly, 11 of 12 of the signature positions are situated within a distance of only 315 nt at the 5’-end region of the 16S rRNA molecule. The signature positions for the phylum planctomyces and relatives are 47
the genus *Planctomyces* (3), Thermotogales (some), Chloroflexus subdiv., Thermus subgr., cyanobacteria (few), *Propionibacterium* gr., *Syntrophomonas* assemblage, *Acetogenium* subgr., *Ach, Shb, Mrc, Acy (Acetobacter)*, *Sbf, Lpp*

G · U: *Planctomyces* subdiv. (few), cyanobacteria (few)

G · A: *Planctomyces* subdiv. (few), alpha-Purple (few)

663 · 742

U · A

A · G

U · A: FCB (some)

665 · 741

G<sup>9</sup> · A<sup>9</sup>

C · G · G · G

G · A: FCB (many), spirochetes (few), cyanobacteria (some), Buc, Dsr, Hlc, Cam (3), *Mycoplasma pneumoniae* gr. (some), Anp, Ast, L (2), Lc (1)

758

A

C, G

A · U: FCB (2), Ric (1), Cam, Bac (1), Fus, Fra, M (3), Dfm (1)

771 · 808

A<sup>9</sup> · U<sup>9</sup>

G · C · U · A

A · U: FCB (3), Spi (1), Bor (1), Acn, Psr, Fus (1), *Arthrobacter* gr. (most), Acy, MWR (1), Mrc (1), MLO (1), Spi (2), Lc (1), L (most), B (2), Cor, ABC, Knc, Armb (1)

822 · 878

C<sup>9</sup> · G<sup>9</sup>

A · U · G · C

C · G: *Thermotogales, Chloroflexus* subdiv. (5), T, FCB (9), Fib, Bs, Trp (2), Bor, Cam (2), *Xanthomonas* gr.

836 · 850

A · U

G · U

A · U: Hal, Flc (1), Fis, Bor, Chr (1), Xyl, Dif, Dfm, Spm, Lc (1)

894 · 905

A · U

G · U

A · U: *M. pneumoniae* group, Pps (1)

941 · 1342

A<sup>9</sup> · U<sup>9</sup>

G · C


1115 · 1185

U<sup>9</sup> · A<sup>9</sup>

C · G · U · G

A · U: Lpp, Bor (most), Hls, t-subdiv. (some), Myb (1), Mic (1), Ppr (1), MWR (9)

1163 · 1173

U · A · C<sup>9</sup>

A · G · C

U · A: t-subdiv., D, green sulfur bacteria and rel., *Sporolactobacillus* subv., MWR (1), FCB (some), (Trp (1), Dsv (1), Bde, T (some), Fusobacteria and rel., Clostridia and rel. (some))

1202

G<sup>9</sup>

U


1253 · 1284

A<sup>9</sup> · U

G · C · G · U

A · U: Got (1), Sng (1), FCB (few), *Leptonema* group, *Ehrlichia* assemblage (most), Leg (1), Msr (1), Clostridia and rel. (some), Cam (few), Fex, Cor (1), Qui, Acp (1), Shb (1)

1352 · 1370

G<sup>9</sup> · C<sup>9</sup>

C · G · U · G

G · C: *TOR, C. thermocellum* subgr.

1356 · 1366

U<sup>9</sup> · A<sup>9</sup>

A · U · G · C

U · A: Lpp, FCB (some), Spi (2), Bor, AltM, *Rhodobacter* gr. (most)

U · G: Cfx (2), Cer (1), Trp (1)

U · U: TOR (some)

---

* Nucleotide positions of bases or base pairs are given according to the *E. coli* numbering (6).

* Nucleotide composition also found in 16S rRNA gene of the *Simkania* sp. strain. Boldface type indicates that the residue is regarded as a signature position for the genus *Chlamydia*.

* Nucleotide composition also found in 16S rRNA gene of *Chlamydia* sp. strain br9 (acc. no. Y07556).

* Residue commonly found among (eu)bacteria.

* Nomenclature and abbreviations are according to the RDP (29). Numbers within parentheses denote the number of sequences in which the residue is found.

Abbreviations: MLO, mycoplasmalike organisms; L, *Lactobacillus*; α, γ, and ε, alpha, gamma, and epsilon subdivisions, respectively, of the purple bacteria; FCB, *Flexibacter-Cytophaga-Bacteroides* phylum; MWR, mollicutes and wall relatives; TOR, thermophilic oxygen reducer; ATS, atypical treponeme subdivision.

(G), 48 (A), 52 (G), 53 (G), 110 (A, except for *Planctomyces limnophilus*), 242 (U, except for *Planctomyces limnophilus*), 284 (G), 331 (U), 353 (U), and 361 (A) as defined by Weisburg et al. (38). Furthermore, the list can be extended by including the positions 359 (C) and 807 (U).

The unique nucleotide positions for the two phylogenetic clusters, the *C. pneumoniae* cluster and the *C. psittaci* cluster, are listed in Table 5. The residues in six of these positions can also be regarded to be signature nucleotides (three positions for each cluster), which further justifies the classification of the genus *Chlamydia* into two distinct clades. It is striking that more than a third of the nucleotide positions unique to the respective cluster are found in a stretch between positions 1006 and 1022 in the locale of variable region V3 of the 16S rRNA molecule (14). The locations of these positions in the primary structure of the 16S rRNA molecule and their nucleotide composition are illustrated in Fig. 2. Secondary structure modeling shows that the actual portion of dense nucleotide information is situated in a stem region which is terminated by a tetraloop starting at position 1013. The nucleotide composition of the loop in this helix follows the main alternative sequence, GAAA, found in the domain *Bacteria* (45) and is conserved in
both clusters. Moreover, the loop is closed by A · U, which is the dominant closing pair for this loop in the 16S rRNA molecule of eubacteria (45). Prediction of the secondary structure of the resulting 16S rRNA molecule from the sequence alignment of Fig. 2 reveals that the juxtaposed stalk differ significantly in G+C content. The G+C content is lower in this region for the members of the C. pneumoniae cluster than those of the C. psittaci cluster. The evolutionary events leading to this change of nucleotide composition in this region probably occurred at about the same time and most probably also occurred once, rather than on several independent occasions. The folding pattern in the actual region remains to be shown, but the amount of canonical and noncanonical base pairs between the two clusters is likely to differ.

The signature positions and the unique nucleotide residues presented in this section not only contribute to the likelihood analysis of branching order within the tree but are also useful in development of PCR-based diagnostic systems which can facilitate the detection of these organisms that are difficult to diagnose by cultivation.

C. pneumoniae cluster. The C. pneumoniae cluster contains the human pathogen, C. pneumoniae and the equine isolate N16 (Fig. 1). The members of this cluster showed tight clustering with a similarity value of ≈98.9% (Table 3). The 16S rRNA sequences of the equine isolate N16 differed only by 1 to 15 nt from those of the human strains of C. pneumoniae. Originally, the N16 isolate was classified as a strain of the species C. psittaci, since iodine staining of inclusion bodies was found to be negative (43). Further characteristics justifying this classification were the morphology of elementary bodies and the plasmid content of N16. Recently, strain N16 was found to be closely related to human isolates of C. pneumoniae in a phylogenetic study based on MOMP gene sequences (39). Nevertheless, their MOMP gene nucleotide sequences differed by 5.5% from each other, which indicated that the horse isolate, N16, was distinct from the human strains; therefore, Storey and coworkers regarded strain N16 as a member of the species C. pneumoniae. Furthermore, a certain cross-reactivity was observed in antigenic tests against four monoclonal antibodies between the N16 strain and the human isolates of C. pneumoniae (39). No cross-reaction was noted against antigens of strains from C. psittaci, C. trachomatis, and C. pecorum. These findings are consistent with the phylogeny inferred from nucleotide sequences of the 16S rRNA molecule presented in this work. An appropriate taxonomic challenge would therefore be to classify the strain N16 into a subspecies of C. pneumoniae (e.g., Chlamydia pneumoniae subsp. equi). However, DNA-DNA reassociation experiments must be performed to support this classification or, eventually, elevate strain N16 to a higher rank. The tree in Fig. 1 therefore represents a maximum likelihood phylogenetic placement of strain N16 forming an early branch of the C. pneumoniae cluster supported by the fact that N16 also shares some common ancestral features with C. psittaci as discussed above.

The partial 16S rRNA sequence of two patient samples (acc. no. I12374 and I12375) and a strain of C. pneumoniae isolated from coronary atheromatous lesion (28) were found to be identical with regard to nucleotide sequence length and composition. However, the 851 nt from these strains differed from the human strains P1 (Parola) and TW-183 in two and three

### Table 5. Unique nucleotide positions in the 16S rRNA molecule correlated to the respective Chlamydia cluster

<table>
<thead>
<tr>
<th>Position</th>
<th>Residue present in:</th>
<th>Comments</th>
</tr>
</thead>
<tbody>
<tr>
<td>135</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>166</td>
<td>C</td>
<td>U</td>
</tr>
<tr>
<td>228</td>
<td>A</td>
<td>A</td>
</tr>
<tr>
<td>234</td>
<td>U</td>
<td>C</td>
</tr>
<tr>
<td>265</td>
<td>G</td>
<td>G</td>
</tr>
<tr>
<td>269</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>278</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>419</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>420</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>637</td>
<td>U</td>
<td>U</td>
</tr>
<tr>
<td>868</td>
<td>C</td>
<td>C</td>
</tr>
<tr>
<td>989</td>
<td>G</td>
<td>A</td>
</tr>
</tbody>
</table>

\( ^{a} \text{Nucleotide positions are according to the E. coli numbering (6).} \)

\( ^{b} \text{Residue present in chlamydiae belonging to the C. pneumoniae cluster with no exceptions.} \)

\( ^{c} \text{Residue present in chlamydiae belonging to the C. psittaci cluster with no exceptions.} \)

\( ^{d} \text{The exceptions are restricted to 16S rRNA sequences of major taxa in which the actual residue is found. Nevertheless, the positions furnished with comments are regarded as nucleotide signatures for the actual Chlamydia cluster, since they are found in \(<5\% \) (usually much less) of all deposited sequences. The nomenclature of phylum, genus, and other phylogenetic groups follows that suggested by the RDP (29). Abbreviations: \( \alpha \)-alpha subdivision of the purple bacteria; FCB, Flexibacter-Cytophaga-Bacteroides phylum.} \)
<table>
<thead>
<tr>
<th>Gene Name</th>
<th>Accession</th>
<th>Chromosome</th>
<th>Location</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>gtnA</td>
<td>X28654</td>
<td>X</td>
<td>09000</td>
<td>Integration</td>
</tr>
<tr>
<td>gtnB</td>
<td>X28655</td>
<td>X</td>
<td>09000</td>
<td>Integration</td>
</tr>
<tr>
<td>gtnC</td>
<td>X28656</td>
<td>X</td>
<td>09000</td>
<td>Integration</td>
</tr>
<tr>
<td>gtnD</td>
<td>X28657</td>
<td>X</td>
<td>09000</td>
<td>Integration</td>
</tr>
</tbody>
</table>

**Legend**
- **gtnA**: Gene encoding the transposase subunit of the gtnB-gtnC-gtnD operon.
- **gtnB**: Gene encoding the gtnB subunit of the gtnB-gtnC-gtnD operon.
- **gtnC**: Gene encoding the gtnC subunit of the gtnB-gtnC-gtnD operon.
- **gtnD**: Gene encoding the gtnD subunit of the gtnB-gtnC-gtnD operon.
positions, respectively. A tree derived from a data set consisting of about 850 nt positions placed the former three strains on the same branch as and close to the human strains (data not shown). Thus, the Chlamydia cluster branches into two lineages: the single strain-containing N16 lineage and the lineage which contains human isolates.

C. psittaci cluster. The C. psittaci cluster is composed of the three species C. psittaci, C. pecorum, and C. trachomatis. The tree constructed with the maximum-likelihood method (Fig. 1) indicated an obvious internal cluster instability, as judged from a bootstrap value of only 53%. As discussed above, a different branching order within this cluster was obtained by the neighbor-joining method. This instability was recognized as a conflict of C. pecorum grouping with either C. trachomatis (suggested by maximum-likelihood and maximum-parsimony methods) or with C. psittaci (suggested by the neighbor-joining method). Therefore, the phylogenetic instability regarding the branching order of the species belonging to this cluster was investigated by performing analysis of nucleotide residues at specific sites. The two alternative grouping patterns were compared first by counting the number of residues that supported each alternative in a parsimony tree analysis and second by counting the number of nucleotide residues in a manual signature analysis as follows. Consensus sequences from the strains of the respective Chlamydia species were compiled, and the nucleotide positions which differed between the four species of the genus Chlamydia were investigated with respect to the branching order of its members. A total of 121 aligned nucleotide positions for which all sequences had data were found to differ among the 16S rRNA sequences of Chlamydia spp. (Fig. 2). Amf, ambiguous positions within the consensus sequences were not used to decide the branching order. The investigated positions were chosen such as being justified from a chlamydial ancestral point of view. Therefore, the branching order was scrutinized by using parsimonious nucleotide residues which resulted in branches of only 1 step in length and which held for both the C. pneumoniae cluster and either C. psittaci, C. pecorum, or C. trachomatis. Therefore, this search for consistency of the actual node involved positions for which only two alternative nucleotide residues were present. In total, there were 12 such positions in the complete alignment, and they are indicated with asterisks in Fig. 2. Of these, there were seven nucleotide positions (i.e., 132, 139, 165, 190, 191, 606, and 632) which supported the maximum-likelihood and maximum-parsimony tree topology and four positions (i.e., 443, 1216, 1244, and 1293) which supported the branching order in the neighbor-joining tree. Only one nucleotide position (i.e., 264) clustered C. psittaci with C. trachomatis. For comparison, when counting the number of nucleotide positions that supported the two different alternatives in the parsimony tree analysis, the grouping suggested by the neighbor-joining method was supported by only two positions while the topology derived by maximum-likelihood and maximum-parsimony methods was supported by five positions. Thus, on the level of individual character states, the results were congruent with the statistical analyses presented above. In conclusion, the tree displayed in Fig. 1 had the most supportive evidence in describing the evolutionary relationships among the members of the genus Chlamydia as inferred from 16S rRNA gene sequences.

Previously, C. trachomatis has been found to harbor two rRNA genes of the small subunit (12). Partial 16S rRNA sequences which consisted of about 550 nt of the U2-to-U5 region (14) were obtained from five human strains of C. trachomatis representing different serovars (data not shown). Strains of the serovars A, C, D, and F shared the same 16S rRNA sequence which was found to be identical to that previously reported for C. trachomatis 434, representing serovar L2. The strain of serogroup L3 was found to differ from strains of serovars A, C, D, and F in two positions in the segment between the U2 and U5 region of the 16S rRNA gene. No polymorphic position was observed for any of the strains in this
part of the 16S rRNA gene. This indicated that *C. trachomatis* is a homogeneous species with respect to nucleotide variation in the 16S rRNA gene, and it is also highly likely that this species exhibits a low variation between strains isolated from humans, as judged from the near-identical 16S rDNA sequences. Sequence information of the 16S rRNA of the *C. trachomatis* serovars was not used for construction of phylogenetic trees, and therefore, these sequences have not been deposited in GenBank.

Analysis of the recently deposited partial sequences from other chlamydial strains within the *C. psittaci* cluster further supported the significance of this phylogenetic entity. Only the strains for which 16S rRNA sequence information of 570 nucleotide positions have been deposited showed any nucleotide difference from the strains of the discussed species of this cluster. Two partial sequences of *C. psittaci* were found to be unique, however, showing few differences from the near-complete sequences of strains 6BC and OEA of this species. The partial 16S rRNA data from *C. psittaci* strains isolated from a guinea pig (U61768) and a pigeon (U61767) showed three differences from each other and ≤3 positions that differed from strains 6BC and OEA of *C. psittaci*. However, despite the variation of the 16S rRNA nucleotide sequences among strains of *C. psittaci*, a tree inferred from a data set consisting of sequences that were 570 nt long showed that the strains of this species grouped tightly together (data not shown).

A striking higher-order structural feature was observed for the species *C. pecorum* in position 748 of the 16S rRNA molecule. The two different 16S rRNA sequences obtained from the *C. pecorum* strains were both found to truncate the molecule at position 748 by one nucleotide, an adenosine. This attribute is rarely seen among the eubacteria, since the residue at position 748 is a highly conserved unilaterally bulged residue in the 16S rRNA molecule of the members of this domain.

A partial 16S rRNA sequence of *C. pecorum* VR 628, which is the same strain as the type strain Bo/E8 investigated in this study, was deposited recently under acc. no. U61770. The sequence which consisted of 570 nt was retrieved and aligned. Comparison with the nucleotide data determined in this study from the 16S rRNA gene of *C. pecorum* Bo/E8 revealed six discrepancies. The deposited sequence of the *C. pecorum* Bo/E8 (VR628) showed 5 nt differences compared with that determined in this study of which two were localized to the highly conserved U2 region of the 16S rRNA molecule (14). Moreover, the higher-order structural feature for *C. pecorum* presented in this work and characterized by the missing adenosine residue in position 748 did not hold for the deposited partial sequence (U61770). Rechecking these nucleotide deviations, totaling 6, between the near-complete sequence obtained in this work and the 570 nt deposited for *C. pecorum* Bo/E8 (VR628) confirmed that our sequence is accurate. Our analysis also included a comparison of the conserved nature of the secondary structure of the 16S rRNA molecule in these positions and examination of the consistency with the rbosomal sequences of the other strains of the genus *Chlamydia*, which supported the veracity of our nucleotide sequences. Surprisingly, there was another 507-nt-long 16S rRNA sequence deposited for *C. psittaci* W73 (acc. no. U61769) which was found to be identical to the partial one of *C. pecorum* Bo/E8 (U61770). Cladograms based on 570 nt positions clearly placed these strains close to the *C. pecorum* strains investigated in this study (tree not shown). In conclusion, we found the deposited data of strain Bo/E8 of *C. pecorum* (U61770) and strain W73 of *C. psittaci* (U61769) somewhat confusing and in need of clarification or confirmation. Therefore, we do not, at this point, believe that the recently deposited partial 16S rRNA sequence undermines the *C. pecorum* data from strains Bo/E8 and BE presented in this work. We also regard the missing adenosine residue in position 748 of the 16S rRNA molecule of *C. pecorum* to be idiosyncratic for this species until otherwise proven.

**Evolution of the genus Chlamydia.** Analysis of nucleotide sequences of the 16S rRNA gene has revolutionized bacterial taxonomy based on phylogeny (31) and studies of microbial evolution (44). Surprisingly little is known about the phylogeny of the genus *Chlamydia*. In this study we have investigated the phylogenetic relationships between the four presently known *Chlamydia* species including some strain variants. Sequence data from the 16S rRNA genes were used to reconstruct the evolutionary history within the genus *Chlamydia* by using different phylogenetic inference methods. The reliability of the suggested tree (Fig. 1) was tested with both statistical analyses and nucleotide signature analysis. Our results suggest that the genus *Chlamydia* can be regarded as representing two distinct phylogenetic clusters, namely, the *C. psittaci* cluster and the *C. pneumoniae* cluster. Both clusters were found to be lines of descendants sharing a common ancestor. Strain N16 shares some of the phenotypical features with those of *C. psittaci* and has a genotype similar to that of *C. pneumoniae*, as discussed above, and the early branching off of *C. psittaci* and strain N16 from their respective cluster are therefore highly likely. Evolutionarily, this indicates that N16 and/or *C. psittaci* carries ancestral features of the originating species of the genus *Chlamydia* since they (i) form early branches of their subgroups and (ii) share biological features such as plasmid content, morphology of elementary bodies, and similar iodine staining patterns of inclusion bodies (43).

**ACKNOWLEDGMENTS**

This work was financially supported in part by grants from the Göran Gustafsson Foundation and the Swedish Engineering Science Council to Mathias Uhlen and Thomas Leitner.

**REFERENCES**


