Conserved indels in essential proteins that are distinctive characteristics of Chlamydiales and provide novel means for their identification

Emma Griffiths,1 Astrid K. Petrich2 and Radhey S. Gupta1

Correspondence
Radhey S. Gupta
gupta@mcmaster.ca

Department of Biochemistry and Biomedical Sciences1 and Department of Pathology and Molecular Medicine2, McMaster University, Hamilton L8N 3Z5, Canada

All known chlamydiae are either proven human or animal pathogens or possess such potential. Due to increasing reports of chlamydiae diversity in the environment, it is important to develop reliable means for identifying and characterizing Chlamydiales species. The identification of environmental chlamydiae at present relies on their branching pattern in 16S rRNA trees, as well as 16S/23S consensus motifs which display variability. At present, no reliable molecular signatures are known which are unique to all Chlamydiales species. Besides the rRNAs, sequence information for different Chlamydiales is not available for any other gene sequence. In this report, a number of molecular signatures are described that consist of conserved inserts and deletions (indels), in widely distributed proteins [RNA polymerase α subunit (RpoA), elongation factor (EF)-Tu, EF-P, DNA gyrase B subunit (GyrB) and lysyl-tRNA synthetase (LysRS)], that are distinctive characteristics of all available chlamydiae homologues (from Chlamydiaceae species and Parachlamydiaceae sp. UWE25) and not found in any other bacteria. Using PCR primers for highly conserved regions in these proteins, the corresponding fragments of these genes from Simkania negevensis, Waddlia chondrophila, and in a number of cases for Neochlamydia hartmanellae, covering all families within the phylum Chlamydiae, have been cloned and sequenced. The shared presence of the identified signatures in these species provides strong evidence that these molecular signatures are distinctive characteristics of the entire order Chlamydiales and can be used to reliably determine the presence of chlamydiae or chlamydia-related organisms in environmental samples. The sequence information for these protein fragments was also used to determine the interrelationships among chlamydiae species. In a phylogenetic tree based on a combined dataset of sequences from RpoA, EF-Tu, EF-P and GyrB, the environmental chlamydiae (i.e. Simkania, Waddlia and Parachlamydia) and the traditional Chlamydiaceae (i.e. Chlamydophila and Chlamydia) formed two distinct clades. Similar relationships were also observed in individual protein phylogenies, as well as in a 16S rRNA tree for the same species. These results provide evidence that the divergence between the traditional Chlamydiaceae species and the other chlamydia families occurred very early in the evolution of this group of bacteria.

INTRODUCTION

The chlamydiae group of species, most of which are important human and animal pathogens, form a distinct phylum within the Bacteria (Fields & Barnes, 1992; Kalayoglu & Byrne, 2001). The phylum Chlamydiae contains only a single class (Chlamydiaceae) and order (Chlamydiales). The order Chlamydiales is currently composed of four families, the Chlamydiaceae, which contains all of the traditional chlamydia species, and three more recently identified families of chlamydia-related organisms: the Parachlamydiaceae (composed of Parachlamydia and Neochlamydia), Waddliaceae and Simkaniaceae (Amann et al., 1997; Birtles et al., 1997; Kahane et al., 1999; Rurangirwa et al., 1999; Horn et al., 2000, 2004; Corsaro et al., 2003). Chlamydia species, until a few years ago, were generally believed to be intracellular parasites of only a limited host range. However, recent studies provide evidence that they are also present in a wide variety of hosts, including fish, insects, turtles and amoebae (Ossewaarde & Meijer, 1999; Everett, 2000; Horn & Wagner, 2001, 2004; Corsaro et al., 2003) (see also www.chlamydiae.com). In recent ecological sampling studies, novel chlamydia-related sequences have also been identified from a wide variety of environments, including
freshwater, soil, sewage and water conduit systems (Horn & Wagner, 2001; Corsaro et al., 2002, 2003; Kahane et al., 2004; Collingro et al., 2005). As a result, it is now recognized that the diversity of chlamydiae, in terms of both their numbers and their clinical involvement, is at present significantly underestimated (Ossewaarde & Meijer, 1999; Fritsche et al., 2000; Corsaro et al., 2003; Thao et al., 2003). The detection of novel chlamydiae-related species (which are generally non-cultur able) in environmental samples is now possible based entirely on the PCR amplification of 16S rRNA sequences and the observed clustering of a sequence in the rRNA trees with those of the known chlamydiae species (Rurangirwa et al., 1999; Everett et al., 1999b; Horn et al., 2000, 2004; Corsaro et al., 2002; Thao et al., 2003). Although consensus sequences for chlamydiae species are found in the 16S and 23S rRNA, these regions display sequence variability. Because of this, and the fact that these motifs do not clearly distinguish chlamydiae species from all other bacteria, assessing the membership of more-divergent species based on these criteria may prove difficult. Since the branching patterns of species in phylogenetic trees are also known to be affected by a large number of variables (Felsenstein, 1988; Gupta, 1998; Moreira & Philippe, 2000; Baldauf, 2003), it is important to develop other reliable markers for identification of chlamydiae species.

The task of identifying novel molecular characteristics that can provide reliable markers for chlamydiae is facilitated by the availability of genomic sequences. The complete genome sequences of five chlamydiae species, Chlamydophila (Chlam.) pneumoniae (four different strains: AR39, CWL029, J139, TW-183) (Read et al., 2000), Chlamydo phila caviae GPIC (Read et al., 2003), Chlamydia (Chl.) trachomatis (serovar D) (Stephens et al., 1998), Chlamydia muridar um (Read et al., 2000) and an environmental chlamydia, Parachlamydia sp. UWE25 (Horn et al., 2004), are now available. The comparative analyses of these genomes provide a powerful means for identifying novel molecular characteristics that are distinctive to different groups of bacteria (Hatch, 1998; Kalman et al., 1999). In an earlier study, we reported a number of conserved inserts and deletions (indels) in widely distributed proteins that were uniquely shared by Chlamydiaceae species (Griffiths & Gupta, 2002). However, most of these indels were small (1 aa), and for most of them sequence information from other chlamydiae families was not available. In the present work, which extends this earlier study, we have identified a number of prominent chlamydiae-specific conserved indels in essential proteins which are found in all bacteria, and in some cases in species from all three domains (Olsen & Woese, 1997). The proteins which contain chlamydiae-specific signatures include RNA polymerase α subunit (RpoA), elongation factor Tu (EF-Tu), DNA gyrase β subunit (GyrB), elongation factor P (EF-P) and llysyl-tRNA synthetase (LyrRS). Sequence information for most of these proteins (RpoA, EF-Tu, EF-P and GyrB) was obtained from species (Simkania, Waddlia and Neochlamydia) covering all of the known chlamydiae families. The presence of these signatures in all of these species provides evidence that they are distinctive characteristics of the phylum Chlamydiales. For the Simkania, Waddlia and Neochlamydia species, apart from the 16S and 23S rRNA, very little sequence information is available in the databases. Hence, we have carried out a phylogenetic analysis based on a concatenated dataset of sequences from these genes (RpoA, GyrB, EF-P and EF-Tu), to determine the interrelationships among different chlamydiae species.

METHODS

Culturing of chlamydiae and DNA extraction. Simkania negevensis (ATCC VR1471), Waddlia chondrophila (ATCC 1470) and Neochlamydia hartmannellae (ATCC 50802) were obtained from ATCC. Sim. negevensis and Wad. chondrophila were propagated in monolayers of HEp2 cells in a similar manner to that described for Chlam. pneumoniae in earlier work (Robinl et al., 1992; Mahony et al., 2000). After incubation of infected cell cultures for 72 h at 37°C in a 95 % air/5 % CO2 atmosphere, the bacteria were harvested by disrupting the HEp2 cells with sterile glass beads, followed by sonication and centrifugation at 250 g for 5 min (Mahony et al., 2000). DNA from these samples was extracted using the blood protocol of the Qiagen DNA Minikit according to the manufacturer’s specifications. DNA from Neo. hartmannellae was directly extracted from the cells obtained from ATCC. The identity of different Chlamydiales was confirmed using 23S rDNA PCR amplification followed by DNA sequence analysis (Everett et al., 1999b). The complete genomes of Chl. trachomatis, Chl. muridarum, Chlam. pneumoniae, Chlam. caviae and Parachlamydia sp. UWE25 are available from the NCBI database (www.ncbi.nlm.nih.gov).

Identification of chlamydiae-specific signatures. Multiple sequence alignments for a large number of proteins have been created in our earlier work (Gupta, 1998, 2000, 2004; Griffiths & Gupta, 2001, 2002, 2004; Gupta et al., 2003) (see also www.bacterialphylogeny.com). To search for chlamydiae-specific signatures, these alignments were visually inspected to identify any indel that was uniquely present in all available chlamydiae homologues and flanked by conserved sequences. The indels which were not flanked by conserved regions, and/or which were not present in all chlamydiae, were omitted from further consideration in this study. The chlamydiae specificity of potentially useful indels was further evaluated by carrying out additional BLAST searches (Altschul et al., 1997) on short sequence segments (usually 60–100 aa) containing the indel and the flanking conserved regions. The purpose of these BLAST searches was to obtain sequence information from all available species to ensure that the identified signatures were only present in the chlamydiae homologues. The sequence information for various useful signatures, which were chosen for further investigation, was compiled into signature files, such as those shown in Figs 1–3.

PCR amplification and sequencing. Oligonucleotide primers, in opposite orientations, were designed for conserved regions in the sequence alignment that flanked the signatures, based on the sequences of the genes/proteins from available chlamydiae species. Degeneracy was incorporated into the primers to account for nucleotide variability at different sites in the alignment. The primers were synthesized at MOBIX, McMaster University. The sequences of various PCR primers that were used to amplify different gene fragments are given in Table 1. PCR was performed in a Techne Tecgene thermocycler under the following conditions. Each reaction had a final volume of 10 μl, and a Mg2+ concentration of 4 mM was used during amplification with all primer sets for each DNA strain tested. Also, 2 % DMSO (v/v, final concentration) was added to the reaction mixtures.
Table 1. Sequences of PCR primers

The unusual bases in these primers are as follows: N=A, T, C or G; Y=T or C; R=A or G; V=G, C or A; W=A or T; S=G or C; H=A, C or T; D=G, A or T; K=G or T.

<table>
<thead>
<tr>
<th>Gene</th>
<th>Primer sequence</th>
<th>Fragment size</th>
</tr>
</thead>
<tbody>
<tr>
<td>EF-Tu, Waddlia</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>375 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
<tr>
<td>EF-Tu, Neochlamydia and Simkania</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>810 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
<tr>
<td>Gyrase B (all chlamydiae)</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>900 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
<tr>
<td>EF-P, Waddlia</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>456 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
<tr>
<td>RpoA, Waddlia</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>522 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
<tr>
<td>RpoA, Simkania</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>477 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
<tr>
<td>LysRS, Waddlia</td>
<td>Forward: 5' GCNCA Run 3'</td>
<td>669 bp</td>
</tr>
<tr>
<td></td>
<td>Reverse: 5' GCNCA Run 3'</td>
<td></td>
</tr>
</tbody>
</table>

Phylogenetic analysis. Multiple sequence alignments of different proteins (RpoA, EF-Tu, GyrB and EF-P) for which sequence information was available from various chlamydiae species were created using the ALIGN PLUS4 program (Scientific and Educational Software). All of these alignments were trimmed to the same length as the amplified fragments. Because the lengths of individual gene fragments were small, to increase the reliability of the phylogenetic analyses, these sequences were concatenated to create a larger alignment consisting of characters from the same groups of species. Any gapped regions, or other regions where low sequence conservation was observed, were removed from this combined alignment. Phylogenetic analysis on the combined dataset was carried out in both the presence and absence of the signature regions to determine their influence on the branching pattern. Genetic distances between bootstrapped datasets (100 times) of aligned sequences were calculated after correcting for multiple amino acid substitutions. Neighbour-joining trees from these distances were constructed and a consensus tree was obtained (Felsenstein, 1988). All of the phylogenetic programs used to construct the tree are part of the TREECON for Windows software package (Van de & De Wachter, 1994). The tree was rooted using the sequence data for Thermus aquaticus. A phylogenetic tree for the same species was also constructed for the full-length 16S rRNA sequences (downloaded from the RDP-II database) (Maidak et al., 2001) for comparison purposes.

RESULTS

Identification and characterization of Chlamydiales-specific indels

Our examination of sequence alignments for different proteins identified a number of prominent conserved inserts and deletions, in widely distributed proteins, that were limited to the available chlamydial homologues. These new chlamydiae-specific signatures included a 15–17 aa insert in the RNA polymerase α subunit (RpoA), a 2 aa insertion and a 2 aa deletion in elongation factor Tu (EF-Tu), a 6 aa insert in DNA gyrase B (GyrB), a 1 aa insert in elongation factor P (EF-P) and a 6 aa insert in lysyl-tRNA synthetase (LysRS). The sequence information for these proteins was only available from the four Chlamydiaceae species and Parachlamydia sp. UWE25, whose genomes have been sequenced (Stephens et al., 1998; Read et al., 2000, 2003; Horn et al., 2004). To determine whether these signatures are distinctive characteristics of chlamydiae, sequence information for these genes/proteins from other chlamydia families (Simkania, Waddlia and Neochlamydia) was required. This was obtained by means of PCR amplification and sequencing of the corresponding fragments. A brief description of these signatures, the proteins in which they are found, and the results of our studies are presented below.

The core subunits of RNA polymerase, α, β and β’, are highly conserved in species from all three domains (Olsen & Woese, 1997). In the α subunit (RpoA), which is encoded...
by the rpoA gene, a 17 aa insert is present in the N-terminal region in various Chlamydiaceae homologues (Fig. 1). A slightly smaller insert of 15 aa is also present in the same position in Parachlamydia sp. UWE23. The insert sequences in all of these species are very similar, indicating that they are of common origin. This insert is not found in any species other than chlamydiae, indicating that it is highly specific for this group. To determine whether this indel is present in Sim. negevensis, Wad. chondrophila and Neo. hartmannellae, PCR amplification of DNA from these species was carried out using the primers indicated in Table 1. For Simkania and Waddlia, this led to specific amplification of 0.5 kb fragments, which were cloned and sequenced. Both species were found to contain the 17 aa insert (sequence information included in Fig. 1), providing evidence that this large insert is a distinctive characteristic of various Chlamydiales. Although these primer sets were not successful in amplifying the target fragments from Neochlamydia under our experimental conditions, it is likely that Neo. hartmannellae will contain the insert, as it has been shown to be part of the genus Parachlamydia.

In the protein synthesis elongation factor-Tu (EF-Tu), which is again a highly conserved protein found in species from all three domains (Hashimoto & Hasegawa, 1996), two different chlamydiae-specific signatures are present in the same region of the protein. One of these signatures consists of a 2 aa insert, whereas the other is a 2 aa deletion, Fig. 1.

---

**Proteobacteria**

1. E. coli
2. P. mirabilis
3. H. influenzae
4. P. aeruginosa
5. X. fastidiosa
6. Ral. salmonicida
7. Bor. pertussis
8. Brk. cepacia
9. Nei. meningitidis
10. Brd. melitensis
11. A. tumefaciens
12. C. crescentus
13. Des. acidocitrica (SIP 012409)
14. Chlam. pneumoniae
15. Chlam. caviae
16. Chl. trachomatis
17. Chl. muridarum
18. S. dysenteriae
19. Wad. chondyphila
20. Para. sp. UWE23
21. Pir. sp.
22. Aqu. gellulose
23. Bac. thetaiotaomicron
24. Cyto. butyricum
25. Ob. sp. ECO 7170
26. Lept. interrogans
27. Sy. sp. ECO 6803
28. Wolin. sp. ECO 7170
29. Cl. walkeri
30. N. phosphatasi
31. T.快讯
32. M. lentum
33. Rbf. longum
34. M. legus
35. Cor. glutaminus
36. Str. cholomtor
37. Tcr. whipplei
38. Dm. xyllic
39. Rcb. xenoplusus
40. Syc. thermophilus
41. L. innocuus
42. Cl. perfringens
43. Syc. pyogenes
44. Bac. halodurans
45. Tcmm. fusca
46. Stc. aurca

---

**Other**

1. Bac. thiooxidans
2. Cyto. hydroxidans
3. Leif. proteus
4. Lep. carlsi
5. Leifi. chondrophila
6. Agrobacterium
7. Aquifex
8. Bacillus
9. Bacteroides
10. Bifidobacterium
11. Bordetella
12. Brucella
13. Burkholderia
14. Caulobacter
15. Chlorobium
16. Chloroflexus
17. Chlamydia
18. Chlamydophila
19. Clostridium
20. Corynebacterium
21. Deinococcus
22. Desulfovibrio
23. Escherichia
24. Herpetobacter
25. Mycobacterium
26. Neisseria
27. Parachlamydia
28. Pasteurella
29. Pirellula
30. Pseudomonas
31. Ralstonia
32. Rubrobacter
33. Simplicibacter
34. Sporobacter
35. Streptomyces
36. Staphylococcus
37. Streptococcus
38. Synechocystis
39. Symbiobacterium
40. Thermotoga
41. Thermobifida
42. Trophyrema
43. Waddlia
44. Xylella

---

**Gram negative Bacteria**

1. Bac. thiooxidans
2. Cyto. hydroxidans
3. Leif. proteus
4. Lep. carlsi
5. Leifi. chondrophila
6. Agrobacterium
7. Aquifex
8. Bacillus
9. Bacteroides
10. Bifidobacterium
11. Bordetella
12. Brucella
13. Burkholderia
14. Caulobacter
15. Chlorobium
16. Chloroflexus
17. Chlamydia
18. Chlamydophila
19. Clostridium
20. Corynebacterium
21. Deinococcus
22. Desulfovibrio
23. Escherichia
24. Herpetobacter
25. Mycobacterium
26. Neisseria
27. Parachlamydia
28. Pasteurella
29. Pirellula
30. Pseudomonas
31. Ralstonia
32. Rubrobacter
33. Simplicibacter
34. Sporobacter
35. Streptomyces
36. Staphylococcus
37. Streptococcus
38. Synechocystis
39. Symbiobacterium
40. Thermotoga
41. Thermobifida
42. Trophyrema
43. Waddlia
44. Xylella

---

**Gram positive Bacteria**

1. Bac. thiooxidans
2. Cyto. hydroxidans
3. Leif. proteus
4. Lep. carlsi
5. Leifi. chondrophila
6. Agrobacterium
7. Aquifex
8. Bacillus
9. Bacteroides
10. Bifidobacterium
11. Bordetella
12. Brucella
13. Burkholderia
14. Caulobacter
15. Chlorobium
16. Chloroflexus
17. Chlamydia
18. Chlamydophila
19. Clostridium
20. Corynebacterium
21. Deinococcus
22. Desulfovibrio
23. Escherichia
24. Herpetobacter
25. Mycobacterium
26. Neisseria
27. Parachlamydia
28. Pasteurella
29. Pirellula
30. Pseudomonas
31. Ralstonia
32. Rubrobacter
33. Simplicibacter
34. Sporobacter
35. Streptomyces
36. Staphylococcus
37. Streptococcus
38. Synechocystis
39. Symbiobacterium
40. Thermotoga
41. Thermobifida
42. Trophyrema
43. Waddlia
44. Xylella

---

**Fig. 1.** Partial alignment of RNA polymerase a subunit (RpoA) homologues showing a 15–17 aa inserted (boxed) that is distinctive for various chlamydiae. The sequence information for this insert from Sim. negevensis and Wad. chondrophila was obtained in the present work. Dashes in all sequence alignments indicate identity to the amino acid on the top line. Numbers on the top line indicate the residue position in the E. coli homologue. Sequence information for only a representative number of species from other bacterial groups is presented here. Abbreviations in genus names are as follows: A, Agrobacterium; Agu., Aquifex; Bac., Bacillus; Bact., Bacteroides; Bil., Bifidobacterium; Bord., Bordetella; Bru., Brucella; Burk., Burkholderia; C., Caulobacter;Cb., Chlorobium; Cl., Chloroflexus; Chl., Chlamydia; Chlam., Chlamydiophila; Clo., Clostridium; Cor., Corynebacterium; Cyto., Cytophaga; D., Deinococcus; Des., Desulfovibrio; E., Escherichia; H., Haemophilus; Leif., Leifsonia; Lep., Leptospira; Li., Listeria; Myc., Mycobacterium; Nei., Neisseria; Para., Parachlamydia; Pas., Pasteurella; Pir., Pirellula; Pse., Pseudomonas; Ral., Ralstonia; Rub., Rubrobacter; Sim., Simkia; Sta., Staphylococcus; Str., Streptomyces; Strep., Streptococcus; Sy., Synechocystis; Sym., Synergibacterium; T., Thermotoga; Therm., Thermobifida; Tro., Tropheryma; Wad., Waddlia; X., Xylella.
Protein signatures specific for Chlamydiales

Proteobacteria

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>E. coli</td>
<td>AAB33326</td>
<td>DQ117555 (-)</td>
</tr>
<tr>
<td>Psa. nortwicis</td>
<td>AAK03341</td>
<td>DQ117556 (-)</td>
</tr>
<tr>
<td>M. influenzae</td>
<td>AAC22932</td>
<td>DQ117557 (-)</td>
</tr>
<tr>
<td>N. meningitidis</td>
<td>AAF05990</td>
<td>DQ117558 (-)</td>
</tr>
<tr>
<td>Rps. anacocorum</td>
<td>CAM7650</td>
<td>DQ117559 (-)</td>
</tr>
<tr>
<td>A. tumefaciens</td>
<td>CAM79345</td>
<td>DQ117560 (-)</td>
</tr>
<tr>
<td>C. crescentus</td>
<td>CAM25161</td>
<td>DQ117561 (-)</td>
</tr>
<tr>
<td>Brev. melitensis</td>
<td>AAL51923</td>
<td>DQ117562 (-)</td>
</tr>
<tr>
<td>Rel. pylori</td>
<td>ACE26711</td>
<td>DQ117563 (-)</td>
</tr>
<tr>
<td>Camp. jejuni</td>
<td>CAM85100</td>
<td>DQ117564 (-)</td>
</tr>
</tbody>
</table>

Chlamydiaciales

Chlamydiales

<table>
<thead>
<tr>
<th>Species</th>
<th>Accession</th>
<th>Sequence Alignment</th>
</tr>
</thead>
<tbody>
<tr>
<td>Wad. chondrophila</td>
<td>AY945401</td>
<td>DQ117565 (-)</td>
</tr>
<tr>
<td>Sim. negevensis</td>
<td>YP903594</td>
<td>DQ117566 (-)</td>
</tr>
<tr>
<td>Neo. hartmanellae</td>
<td>AY945408</td>
<td>DQ117567 (-)</td>
</tr>
<tr>
<td>Aqu. acidocaldarius</td>
<td>AAY31519</td>
<td>DQ117568 (-)</td>
</tr>
<tr>
<td>Fib. succinogenes</td>
<td>AAX41394</td>
<td>DQ117569 (-)</td>
</tr>
<tr>
<td>Bact. forsythus</td>
<td>AAB8139</td>
<td>DQ117570 (-)</td>
</tr>
<tr>
<td>Por. gingivalis</td>
<td>AAA8134</td>
<td>DQ117571 (-)</td>
</tr>
<tr>
<td>Co. vibrioforme</td>
<td>AAM54322</td>
<td>DQ117572 (-)</td>
</tr>
<tr>
<td>Flir. morini</td>
<td>AAGU7903</td>
<td>DQ117573 (-)</td>
</tr>
<tr>
<td>Tre. pallidu</td>
<td>AAI6172</td>
<td>DQ117574 (-)</td>
</tr>
<tr>
<td>Sp.</td>
<td>AAI3132</td>
<td>DQ117575 (-)</td>
</tr>
<tr>
<td>Nostoc sp. PCC 7120</td>
<td>CAM86036</td>
<td>DQ117576 (-)</td>
</tr>
<tr>
<td>Cf. auranticus</td>
<td>CAM4193</td>
<td>DQ117577 (-)</td>
</tr>
<tr>
<td>D. radiodurans</td>
<td>ATT1600</td>
<td>DQ117578 (-)</td>
</tr>
<tr>
<td>Th. aquaticus</td>
<td>CAM6499</td>
<td>DQ117579 (-)</td>
</tr>
<tr>
<td>1. maritima</td>
<td>AAB72315</td>
<td>DQ117580 (-)</td>
</tr>
<tr>
<td>Sl. coelenter</td>
<td>999152</td>
<td>DQ117581 (-)</td>
</tr>
<tr>
<td>Brev. japonicus</td>
<td>CAM4490</td>
<td>DQ117582 (-)</td>
</tr>
<tr>
<td>Cor. glutamicum</td>
<td>CAM45323</td>
<td>DQ117583 (-)</td>
</tr>
<tr>
<td>K. pneumoniae</td>
<td>AAB95325</td>
<td>DQ117584 (-)</td>
</tr>
<tr>
<td>Clostridium sp.</td>
<td>CAM92127</td>
<td>DQ117585 (-)</td>
</tr>
<tr>
<td>Lys. innocua</td>
<td>CAM6026</td>
<td>DQ117586 (-)</td>
</tr>
<tr>
<td>Bac. subtilis</td>
<td>CAM81889</td>
<td>DQ117587 (-)</td>
</tr>
<tr>
<td>Strept. pyogenes</td>
<td>AAK35386</td>
<td>DQ117588 (-)</td>
</tr>
<tr>
<td>Tre. aurum</td>
<td>AAM57170</td>
<td>DQ117589 (-)</td>
</tr>
</tbody>
</table>

Fig. 2. Partial sequence alignment of EF-Tu sequences, showing two different signatures (an insert and a deletion, boxed) that are distinctive of the chlamydiae species. Sequence information for Wad. chondrophila, Sim. negevensis and Neo. hartmanellae was obtained in this work. Additional abbreviations: Brev., Brevibacterium; Camp., Campylobacter; Fib., Fibrobacter; Hel., Helicobacter; M., Mycoplasma; Neo., Neochlamydia; Por., Porphyromonas; The., Thermus; Tre., Treponema.

DNA gyrase is a type II topoisomerase, which makes transient double strand breaks in DNA to allow another DNA strand to pass through (Levine et al., 1998). DNA gyrase function is essential in bacteria for maintaining the appropriate levels of supercoiling in the chromosome and for transcription and replication processes. The enzyme consists of two subunits, A and B, that combine to form a functional A2B2 complex. The B subunit of DNA gyrase (GyrB) contains a 6 aa insert that is uniquely present in various available chlamydia homologues (Fig. 3). The specificity and distinctiveness of this insert was again tested by PCR amplification of 0.9 kb fragments of gyrB from Wad.

chondrophila, Sim. negevensis and Neo. hartmanellae. The cloning and sequencing of these fragments revealed that the indicated signature insert was present in all three species (see Fig. 3), confirming that it is a reliable molecular marker for all chlamydiae.

We have previously described a 1 aa insert in EF-P (Griffiths & Gupta, 2002) that is found in all Chlamydiaceae species, including Chlamydomphila abortus and Chlamydia suis, which were sequenced in our previous work, as well as Sim. negevensis. This insert is also present in Parachlamydia sp. UWE25, and its sequence is completely conserved in all species. Further work that we have carried out indicates that this insert is also present in Wad. chondrophila (Fig. 4), indicating that in a similar manner to the inserts in RpoA, EF-Tu and GyrB, this insert is also a distinctive characteristic of various chlamydia species.

Another signature that is specific for various chlamydiae was identified in the enzyme lysyl-tRNA synthetase, which catalyses the correct attachment of lysine to its cognate tRNA in the protein synthesis process (Woese et al., 2000). The chlamydial LysRS proteins (from the known

http://mic.sgmjournals.org 2651

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11
On: Mon, 22 Jul 2019 08:49:09

Downloaded from www.microbiologyresearch.org by IP: 54.70.40.11
On: Mon, 22 Jul 2019 08:49:09
Chlamydiaceae and Parachlamydia) were found to contain a 6 aa insert which is unique to this group of species (Fig. 5). Using the PCR primers described in Table 1, we were also successful in amplifying a 0.6 kb LysRS fragment from Wad. chondrophila that was found to contain the indel. However, our attempts to PCR-amplify the corresponding fragment from Sim. negevensis and Neo. hartmanellae using this set of primers were unsuccessful, indicating that the primer region may have undergone mutational changes in these species. Since species representing other chlamydiae families contain this insert, it is likely that this indel will also be found in Simkania and Neochlamydia.

Phylogenetic analysis of chlamydiae species based on concatenated protein sequences

Our understanding of the evolutionary relationships among chlamydiae is based solely on 16S and 23S rRNA sequences (Everett et al., 1999a; Bush & Everett, 2001; Corsaro et al., 2003; Horn et al., 2004). The sequence information for different Chlamydiales species is currently not available for any protein-coding genes. Since in our work we have sequenced fragments for a number of protein-coding genes from different chlamydiae families, it was of interest to carry out a phylogenetic analysis based on these protein sequences and compare it with the 16S rRNA tree. As the amplified fragments generated in this work were small, to obtain a reliable phylogenetic tree we have combined (i.e. concatenated) sequence information for GyrB, RpoA, EF-Tu and EF-P proteins from Wad. chondrophila, Sim. negevensis, Parachlamydia sp. UWE25, Chl. trachomatis, Chl. murrayi, Chlam. pneumoniae, Chlam. caviae and Thermus aquaticus. This concatenated dataset containing 529 aa positions was used to construct a neighbour-joined tree which was bootstrapped (100 times) for statistical analysis (Felsenstein, 1988). Thermus aquaticus was used as an outgroup.
outgroup to root the tree (Fig. 6A). The resulting topology showed that the Chlamydiae appear to have evolved in two distinct groups, the *Chlamydiaceae* and the chlamydiae-like organisms. Within the *Chlamydiaceae*, the genus *Chlamydia* appears to have diverged later than *Chlamydophila*. Among the chlamydiae-like species, *Simkania* branches the earliest, followed by a bifurcating clade including *Waddlia* and *Parachlamydia*. All branches were resolved with significant bootstrap scores, and the observed branching pattern was not affected by the presence or absence of the signature inserts (not shown). These findings were also corroborated by individual protein phylogenies based on these protein sequences (data not shown). It should be mentioned that the amino acid sequences of the insert in the EF-Tu protein (Fig. 2) also distinguish the clades consisting of *Chlamydia* (insert sequence SE), *Chlamydophilina* (insert sequence SQ) and *Waddlia/Parachlamydia* (insert sequence GE). We also constructed a neighbour-joined phylogenetic tree based on full-length 16S rRNA sequences (Maidak et al., 2001) (Fig. 6B). This tree showed a similar topology to that seen in the concatenated protein tree, and branches separating the chlamydiae-like species from the *Chlamydiaceae* were resolved with high statistical support (bootstraps > 90 %).

**Proteobacteria**

**Chlamydiaceae**

- *E. coli* 729046
- *Brucella melitensis* AAL51508
- *Pseudomonas putida* AAK02184
- *Y. pestis* CAE06121
- *H. influenzae* Rd 1369460
- *V. cholerae* AAF95808
- *X. fastidiosa* AAF65271
- *F. aeruginosa* AAG06239
- *Z. mobilis* ZM1 892375
- *Neisseria meningitidis* CAB64535
- *Ca. crescentus* AAK2706
- *A. tumefaciens* AAG09299
- *Sin. neillii* CAC1714
- *R. prowazekii* NS2277
- *Comp. jejuni* CAB7187
- *Hel. pylori* 26695
- *Geo. sulfurreducens* NP 952802
- *Desul. vulgaris* YP 010884
- *Chlam. pneumoniae* 437215
- *Chlam. caviae* NP 829736
- *Chlam. abortus* AT083888
- *Chl. naudu* AAF59011
- *Chl. trachomatis* 319291
- *Chl. suis* AT08569
- *Par. sp. WUE2* YP 008528
- *Wad. chondrophila* AT845494
- *Sin. negevensis* AAE8969
- *Bact. fragilis* AACC2628
- *Port. gingivalis* NP 90486
- *Aqu. aeolicus* 393165
- *Tre. pallidum* 332817
- *Eor. burgdorferi* 393584
- *Sy. sp. PCC6803* 2494276
- *Syn. PCC7942* 2494275
- *Anab. sp. Wad. chondrophila* 2494269
- *Nostoc sp. PCC 7120* BAB76757
- *D. radiodurans* AAF09709
- *T. maritima* CQX298
- *Myc. leprosum* CAC50030
- *Cor. glutamicum* 2494271
- *Str. coelicolor* CABS3731
- *Bac. subtilis* 100397
- *M. pneumoniae* AAB5573
- *L. lactis* AAK04790
- *Lls. innocua* CAC96623
- *Strep. pneumoniae* AAK9199
- *Clo. acetobutylicum* NC_048313
- *Sta. aureus* BAB4854

**Other Gram negative Bacteria**

- *K. pneumoniae* AAD0473
- *R. vinosogenes* AAT929
- *E. coli* 729046
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473
- *K. pneumoniae* AAD0473

**Gram positive Bacteria**

- *Strep. pyogenes* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001
- *S. pneumoniae* AAB001

![Fig. 4. Partial sequence alignment for EF-P homologues showing a completely conserved 1 aa (leucine) insertion that is present in all Chlamydiales species. The sequence for *Wad. chondrophila* was obtained in the present work. Additional abbreviations: Anab., *Anabaena*; L., *Lactococcus*; Sin., *Sinorhizobium*; Z., *Zymomobilis*.](http://mic.sgmjournals.org)
DISCUSSION

Our understanding of the chlamydiae has expanded greatly in recent years. Until a few years ago, the phylum Chlamydiae consisted of a single family, the Chlamydiaceae, which contained a single genus with four species: *Chl. trachomatis*, *Chlamydia pneumoniae* and *Chlamydo- phila psittaci* (Fields & Barnes, 1992; Schachter & Stamm, 1999; Kalayoglu & Byrne, 2001). The host organisms for these intracellular parasites were believed to be mainly vertebrates, with only a few exceptions. However, in 1999 a major revision to the chlamydiae taxonomy was proposed by Everett et al. (1999a) to account for a great deal of new information that had accumulated in the preceding years.

The revised proposal divided the existing Chlamydiaceae family into two genera, *Chlamydia* and *Chlamydophila*, which now contained a total of nine species (Everett et al., 1999a). However, a more significant aspect of this revision was the recognition of three additional families.
of chlamydia-related organisms, the *Waddliaceae*, *Parachlamydiaceae* and *Simkaniaceae*, whose host range varies from humans, rodents, birds, marsupials, insects and turtles to amoeba (as an endosymbiont) (Everett, 2000; Horn & Wagner, 2004). The chlamydiae-related species have now also been detected in numerous other species, including hydra, bivalves, isopod crustacea, spiders, reptiles and fish (Corsaro *et al.*, 2003; Draghi *et al.*, 2004; Kostanjsek *et al.*, 2004; Collingro *et al.*, 2005). Another major development that has taken place in recent years is the finding of chlamydiae-related sequences in a wide variety of environments, including freshwater ponds, soil, sewage and water conduit systems (Corsaro *et al.*, 2002, 2003). As a result of these findings, it is now widely recognized that the diversity of chlamydiae, in terms of both their numbers and their clinical involvement, is currently significantly underestimated (Ossewaarde & Meijer, 1999; Fritsche *et al.*, 2000; Horn & Wagner, 2001; Corsaro *et al.*, 2003; Thao *et al.*, 2003). Because of the zoonotic potential of environmental chlamydiae to humans and animals (Greub & Raoult, 2002; Corsaro & Venditti, 2004; Horn & Wagner, 2004), it is important to develop reliable means for identifying chlamydiae species in different hosts and environments.

All well-studied chlamydiae species are characterized by a unique two-stage life cycle (Fields & Barnes, 1992; Kalayoglu & Byrne, 2001). However, for most new reports of chlamydiae-like species (sequences), which are from non-culturable sources, it is difficult to obtain information in this regard. The primary means used for identifying novel chlamydiae species in different hosts and environmental samples involves PCR amplification using primers based on 16S rRNA sequences, and the clustering of any amplified sequence with the known chlamydiae species in the rRNA trees (Rurangirwa *et al.*, 1999; Everett *et al.*, 1999b; Horn *et al.*, 2000, 2004; Corsaro *et al.*, 2002; Thao *et al.*, 2003). Although this method has proved very useful in advancing our understanding of chlamydial diversity, the branching patterns in phylogenetic trees are known to be affected by a large number of factors (Felsenstein, 1988; Gupta, 1998; Moreira & Philippe, 2000; Baldauf, 2003). Other than their branching pattern in the tree, no distinctive molecular signature for chlamydiae is present in the 16S rRNA sequences. Hence, the availability of other specific and reliable molecular markers for confirming the presence of chlamydiae species in environmental samples should be very useful.

In the present work, we have described several distinctive molecular markers for chlamydiae species in highly conserved proteins. These markers consist of prominent inserts and deletions in highly conserved regions of a number of important proteins (*RpoA, EF-Tu, GyrB, EF-P* and *LysRS*) that are unique to various chlamydiae homologues and not found in any other bacteria. The sequence information for most of these proteins (except *LysRS*) has been obtained from *Waddlia* and *Simkania*, and for *EF-Tu* and *GyrB* from *Neochlamydia* as well. The shared presence of these signatures in all of the chlamydiae families strongly indicates that these signatures were introduced in a common ancestor of the Chlamydiales and that they are distinctive characteristics of the entire phylum. The proteins in which these chlamydiae-specific signatures are present are all involved in essential functions related to information-transfer processes. Most of these proteins are ubiquitous in bacteria, and some (*RpoA, EF-Tu* and *LysRS*) are present in species from all three domains. Because of their essential functions, the primary structures of these proteins are highly conserved. The PCR primers for *EF-Tu* and *gyrase B* used in our work were successful in amplifying the corresponding gene fragments from *Sim. negevensis*, *Wad. chondrophila* and *Neo. hartmanellae*. Because these primers are based on sequence information for *Chlamydiaceae* and *Parachlamydia* sp. UWE25, they are also expected to work for these species. The PCR primers for *RpoA* and *EF-P* also successfully amplified the corresponding fragments from *Sim. negevensis* and *Wad. chondrophila*, but they were not successful with *Neo. hartmanellae* under the experimental conditions used in the present study. However, based on the fact that these sequences are highly conserved, it should be possible to design other primers that would work in cases where PCR amplification was not successful in the present study. These signatures and the PCR primers provide novel means, in addition to the 16S rRNA-based primers, for identifying and confirming the presence of chlamydiae-related sequences in environmental samples. Because of the chlamydiae specificities of these signatures, if any of the amplified sequences are found to contain these signatures, this will provide reliable evidence for the presence of a chlamydiae-like organism in a given sample.

Phylogenetic analysis based on a combined dataset of protein fragments from *RpoA, EF-Tu, EF-P* and *GyrB* proteins indicates that the environmental chlamydiae (*Simkania, Waddlia* and *Parachlamydia*) and the traditional *Chlamydiaceae* (*Chlamydomphila* and *Chlamydia*) species form distinct clades in the resulting tree. Similar relationships were also noted in individual protein phylogenies as well as in a tree for these species based on the 16S rRNA sequences. Thus, it is highly likely that the chlamydiae-like species (*Waddlia, Parachlamydia* and *Simkania*) have diverged from the traditional *Chlamydiaceae* species at a very early stage in the evolution of this group of bacteria. If this is the case, then the chlamydiae-like species may contain a very different range of metabolic, infective and virulence capabilities, as evidenced by the large difference in genome size between *Parachlamydia* sp. UWE25 and all other sequenced *Chlamydiaceae* (Hatch, 1998; Kalman *et al.*, 1999; Read *et al.*, 2000, 2003; Horn *et al.*, 2004).

**ACKNOWLEDGEMENTS**

The work was supported by a research grant form the Canadian Institute of Health Research to R. S. G.
REFERENCES


Griffiths, E. & Gupta, R. S. (2001). The use of signature sequences in different proteins to determine the relative branching order of bacterial divisions: evidence that Fibrobacter diverged at a similar time to Chlamydia and the Cytophaga-Flavobacterium-Bacteroides division. Microbiology 147, 2611–2622.


