Phylogeny and shared conserved inserts in proteins provide evidence that *Verrucomicrobia* are the closest known free-living relatives of chlamydiae

Emma Griffiths and Radhey S. Gupta

Department of Biochemistry and Biomedical Sciences, 1200 Main Street West, Hamilton, L8N 3Z5, Canada

The evolutionary relationships of *Chlamydiales*, *Verrucomicrobia* and *Planctomycetes* were studied based on phylogenetic trees for a concatenated dataset of 11 widely distributed proteins, as well as conserved inserts in several proteins. In phylogenetic trees, a close relationship of chlamydiae to *Verrucomicrobium* was supported by different phylogenetic methods. Although the *Planctomycetes* branched close to the chlamydiae-*Verrucomicrobium* clade, their specific affiliation to these groups was generally not supported. Results are also presented for two conserved inserts, a 6 aa insert in the lysyl-IRNA synthetase and a 3 aa insert in the RNA polymerase β subunit (RpoB), that are uniquely shared by *Verrucomicrobium spinosum* and all available *Chlamydiales* homologues, but which are not found in any of the available *Planctomycetes* or other bacterial homologues. Signature sequences in a number of other proteins [including a large insert (>150 aa) in DNA gyrase B] provide information regarding the branching position of these groups relative to other bacterial phyla. A close and specific relationship of *V. spinosum* to the *Chlamydiales* species, seen both in phylogenetic trees and by means of uniquely shared inserts in protein sequences, strongly indicates that these two groups of species shared a common ancestor exclusive of all other known bacteria. These results suggest that *Verrucomicrobia* may be the closest free-living relatives of the parasitic chlamydiae.

**INTRODUCTION**

The *Chlamydiales*, *Verrucomicrobia* and *Planctomycetes* are presently recognized as three main phyla within Bacteria (Everett *et al.*, 1999; Garrity *et al.*, 2005; Schelsner *et al.*, 2006; Ward *et al.*, 2006). Of these, all known chlamydiae species are obligate intracellular parasites of eukaryotic hosts and they are responsible for a wide spectrum of diseases in humans and animals (Everett *et al.*, 1999; Corsaro & Greub, 2006). The verrucomicrobia and planctomycetes species are found in a wide variety of terrestrial and aquatic environments (Strous *et al.*, 2006; Schelsner *et al.*, 2006; Ward *et al.*, 2006). Although most of them are free-living bacteria, several live in close association with eukaryotic hosts, including some verrucomicrobia species that are endosymbionts of nematodes and ciliates (Wagner & Horn, 2006; Schelsner *et al.*, 2006; Ward *et al.*, 2006). Of these three groups, chlamydiae and planctomycetes, unlike most other bacteria, contain no detectable peptidoglycan in their cell envelopes (Konig *et al.*, 1984; Fox *et al.*, 1990; Ward *et al.*, 2006). The evolutionary relationships among these three lineages are presently not resolved (Ward *et al.*, 2000; Jenkins & Fuerst, 2001; Teeling *et al.*, 2004; Ciccarelli *et al.*, 2006; Wagner & Horn, 2006; Strous *et al.*, 2006; Schelsner *et al.*, 2006; Ward *et al.*, 2006). Wagner & Horn (2006) reported a monophyletic grouping of these three lineages in the 16S rRNA trees, but the bootstrap support of the combined clade was not high. A grouping of the *Planctomycetes* with chlamydiae was also seen in phylogenetic trees based on concatenated sequences for several proteins (Teeling *et al.*, 2004; Strous *et al.*, 2006), but these trees lacked any verrucomicrobia species. In contrast, other studies based on either 16S and 23S rRNA (Ward *et al.*, 2000), or sequences for individual or concatenated proteins (Jenkins & Fuerst, 2001; Ciccarelli *et al.*, 2006), have found no statistical support for a specific relationship between the *Verrucomicrobia, Planctomycetes* and *Chlamydiales* lineages. Hence, it is important to examine the evolutionary relationships among these three groups using other approaches.

In recent years, sequence information for numerous genomes including many chlamydiae (Stephens *et al.*, 2006)
1998; Kalman et al., 1999; Read et al., 2000, 2003; Horn et al., 2004), as well as two planctomycetes (Glockner et al., 2003; Strous et al., 2006), has become available (http://www.ncbi.nlm.nih.gov/genomes/lproks.cgi). In addition, sequencing of genomes from several other species from these groups including Simkania negevensis (Chlamydiidae), Blastopirellula marina and Gemmata obscuriglobus (Planctomycetes) and Verrucomicrobiurn spinosus is under way, and sequence information for many genes/proteins from these species is available at The Institute for Genomic Research web site (http://www.tigr.org). The resulting vast expansion in sequence information makes it now possible to examine the evolutionary relationship among these bacteria by both traditional phylogenetic methods and novel approaches. The shared presence of rare genomic changes (RGCs) such as conserved inserts or deletions (i.e. indels) in protein sequences provides a powerful tool for evolutionary relationships among distantly related groups (Gupta & Griffiths, 2006). In earlier work, we have described many conserved indels in protein sequences that are either specific for the chlamydiae species or provide information regarding their branching relative to other bacterial phyla (Griffiths & Gupta, 2001; Griffiths et al., 2005; Gupta & Griffiths, 2006). In this work, we have examined the evolutionary relationships among chlamydiae, Verrucomicrobiurn and Planctomycetes by traditional phylogenetic methods as well as by identifying conserved indels in protein sequences that are uniquely shared by species from these groups.

RESULTS

Phylogenetic analyses based on concatenated protein sequences

Phylogenetic trees for the bacterial groups were constructed based on concatenated sequences for 11 conserved proteins that are involved in a broad range of functions. The trees were constructed by different methods, and the ML tree is shown in Fig. 1. In the trees obtained using different methods, the species corresponding to various main phyla within Bacteria (e.g. Firmicutes, Actinobacteria, Cyanobacteria, Spirochaetes, Deinococcus-Thermus, Bacteroidetes-Chlorobi, Proteobacteria, etc.) formed distinct clades with high bootstrap scores, but their relative branching was not resolved. All chlamydiae species formed a strongly supported clade in these trees and V. spinosus formed its outgroup with high statistical support (>90 % bootstrap score) by different methods. The Planctomycetes species showed no specific association with the chlamydiae-Verrucomicrobiurn clade in the ML tree (Fig. 1), although in the NJ tree they formed the immediate outgroup of the chlamydiae-Verrucomicrobiurn clade with high (97 %) bootstrap score (results not shown). A similar relationship of the Planctomycetes to the chlamydiae-Verrucomicrobiurn was seen in the MP tree, although the bootstrap score was only 55 % (results not shown).

Conserved inserts in LysRS and RpoB proteins that are unique to the chlamydiae and Verrucomicrobiurn

The RGCs such as conserved inserts and deletions (i.e. indels or signature sequences) in gene/protein sequences that are uniquely shared by particular groups of species provide valuable markers for taxonomic and evolutionary studies (Gupta & Griffiths, 2006). We recently described a number of conserved indels in proteins such as RpoA, EF-Tu, EF-P, gyrase B and LysRS, that based upon the available information were distinctive characteristics of the Chlamydiidae species (Griffiths et al., 2005). Due to lack of sequence information, the presence of these signatures in Verrucomicrobiurn was not determined. To determine whether any of these indels are also present in V. spinosus and various available Planctomycetales species, BLAST searches on these proteins were carried out against the genomes of these species to obtain sequence information for the corresponding homologues. These studies have revealed that the indels in RpoA, EF-Tu, EF-P and gyrase B are specific for the Chlamydiidae and were not found in either V. spinosus or any of the planctomycetes species (viz. Rhodopirellula baltica, B. marina, G. obscuriglobus and also Candidatus Kuenenia stuttgartiensis) (results not shown). However, the 6 aa insert in LysRS, which was previously indicated to be Chlamydiidae-specific, was also present in the V. spinosus homologue in the same position (Fig. 2), although it was not present in any of the

METHODS

Phylogenetic tree construction and analysis of conserved inserts. The sequences for 11 proteins [viz. arginyl-tRNA synthetase, alanyl-tRNA synthetase (AlaRS), elongation factors (EF)-Tu and EF-G, DNA gyrase A and B subunits, RNA polymerase α, β and β’ subunits (RpoA, RpoB and RpoC, respectively), SecA and SecY] for 33 species covering all main phyla of bacteria (e.g. Firmicutes, Actinobacteria, Cyanobacteria, Spirochaetes, Deinococcus-Thermus, Bacteroidetes-Chlorobi, Proteobacteria, etc.) were retrieved from the NCBI database. Sequence alignments for individual proteins were created using the CLUSTAL X (1.83) program using the default settings and then concatenated into a large dataset. The poorly aligned regions from this alignment were removed using the Gblocks 0.91 program (Castresana, 2000). The final alignment file, which contained 5093 positions, is available as Supplementary Data with online version of this paper. Bootstrapped phylogenetic trees for this alignment were constructed by neighbour-joining (NJ), maximum-likelihood (ML) and maximum-parsimony (MP) methods, as described in our recent work (Gupta & Sneath, 2006).

The conserved insert in RpoB that is specific for the chlamydiae and Verrucomicrobiurn was identified by visual inspection of different sequence alignments (Griffiths et al., 2005). The insert in the lysyl-tRNA synthetase (LysRS) was previously identified as a Chlamydiidae-specific signature (Griffiths et al., 2005). The species distribution of various conserved inserts was determined by carrying out BLASTP searches against the NCBI non-redundant (nr) database on a short segment of the sequence (between 80 and 120 aa) containing the insert and its flanking regions (Griffiths et al., 2005). The preliminary sequence data for V. spinosus, B. marina and G. obscuriglobus were obtained by BLAST searches against the partial genomes of these species at The Institute for Genomic Research web site (http://www.tigr.org).
planctomycetes species. For this insert, while the sequence on the left is highly conserved, the sequence downstream to that shown here is not conserved in distantly related species (not shown). However, there is a reasonable degree of conservation on both sides of the insert to ascertain that it is not present in other species.

In addition to the insert in LysRS, we have also identified a novel 3 aa insert in the RpoB protein that is a distinctive characteristic of various Chlamydiales and V. spinosum homologues. Similar to the LysRS insert, this insert is also not found in the RpoB homologues from any other bacteria, including various Planctomycetes (Fig. 3). The sequence of this insert (i.e. RRK) is completely conserved in all Chlamydiales and it is very similar to that found in V. spinosum (RRR). The RpoB homologues from most bacteria, including chlamydiae and V. spinosum, contain another RRR or RKR motif adjacent to this insert. Hence, this insert has very likely resulted from a duplication of this conserved motif in the in Chlamydiales-Verrucomicrobiurn.

The RR(K)R motif in RpoB lies downstream of a region that is involved in binding of the transcription repair protein.
### Common ancestry of chlamydiae and Verrucomicrobium

**Fig. 2.** A 6 aa insert in the lysyl tRNA synthetase that is uniquely shared by various *Chlamydiales* homologues and *V. spinosum*, but not found in any other bacteria. The dashes indicate identity with the amino acid on the top line. The numbers on the top indicate its position in *E. coli* sequence (top line). Accession numbers for various proteins are indicated in the second column.

**Abbreviations in the species names are as follows:**
- **A.,** *Agrobacterium*
- **Aqu.,** *Aquifex*
- **Bac.,** *Bacillus*
- **Bact.,** *Bacteroides*
- **Bde.,** *Bdellovibrio*
- **Bif.,** *Bifidobacterium*
- **Bla.,** *Blastopirellula*
- **Camp.,** *Campylobacter*
- **Cb.,** *Chlorobium*
- **Chlam.,** *Chlamydia*
- **Chlam.,** *Chlamydophila*
- **Clo.,** *Clostridium*
- **Cor.,** *Corynebacterium*
- **Cyto.,** *Cytophaga*
- **D.,** *Deinococcus*
- **E.,** *Escherichia*
- **Gemm.,** *Gemmata*
- **Gloe.,** *Gloeobacter*
- **Hel.,** *Helicobacter*
- **K.,** *Kuenenia*
- **Lep.,** *Leptospira*
- **Myc.,** *Mycobacterium*
- **Nei.,** *Neisseria*
- **Pas.,** *Pasteurella*
- **Per.,** *Persephonella*
- **Por.,** *Porphyromonas*
- **Proto.,** *Protochlamydia*
- **Pse.,** *Pseudomonas*
- **Ral.,** *Ralstonia*
- **Rho.,** *Rhodopirellula*
- **Sim.,** *Simkania*
- **Sta.,** *Staphylococcus*
- **Str.,** *Streptomyces*
- **Strep.,** *Streptococcus*
- **T.,** *Thermotoga*
- **The.,** *Thermus*
- **Tre.,** *Treponema*
- **Ver.,** *Verrucomicrobium*
- **V.,** *Vibrio*
- **Wad.,** *Waddlia*

In earlier work, a number of conserved indels in important housekeeping proteins (e.g. Hsp70, Hsp60, AlaRS, RpoB, RpoC, inorganic pyrophosphatase, CTP synthetase) have been described whose species distribution profiles provide information regarding the relative branching order of bacterial phyla (Griffiths & Gupta, 2001, 2004). These studies have suggested the placement of *Chlamydiae* species in the same position as the *Fibrobacteres-Chlorobi-Bacteroidetes* (FCB) groups. The bacterial phyla consisting of *Firmicutes*, *Actinobacteria*, *Deinococcus-Thermus*, *Cyanobacteria* and *Spirochaetes* were inferred to branch prior to the *Chlamydiae-FCB* groups, whereas *Aquificales* and different divisions of *Proteobacteria* were indicated as late branching groups (Griffiths & Gupta, 2001, 2004). Due to lack of sequence information for *Verrucomicrobia* and *Planctomycetes*, the branching positions of these groups were not inferred. We have examined the presence and absence of various main-line indels in the above proteins in the *Planctomycetes* and *Verrucibacterium* species. For the described indels in all of these proteins, except AlaRS, *Planctomycetes* and *V. spinosum* showed similar distribution profiles, as reported in earlier work for the
chlamydiae species (Griffiths & Gupta, 2001, 2004) (results not shown). Whereas the insert in AlaRS was present in all chlamydiae, V. spinosum and G. obscuriglobus, it was absent from other planctomycetes species. The observed heterogeneity in the distribution profile of this indel could be a result of lateral gene transfer. Overall, the species distribution profiles of these indels provide evidence that Verrucomicrobia as well as Planctomycetes branch at a similar position as the Chlamydiae-FCB groups of bacteria.

A prominent insert in the DNA gyrase B protein now provides clarification regarding the branching position of the Planctomycetes, Chlamydiales and Verrucomicrobia relative to the FCB groups of bacteria. This indel was earlier described in relation to its effects on gyrase function, but its species distribution profile or evolutionary implications were not discussed (Chatterji et al., 2000). In this highly conserved protein, which plays an essential role in the transcription and replication processes in bacteria, a large insert (>150 aa) in a conserved region is uniquely present in all Proteobacteria, Aquificales, different Chlamydiales, as well as in V. spinosum and all four Planctomycetes species (see Supplementary Fig. S1, available with the online version of this paper). However, this insert is not present in the gyrase B homologues from other bacterial phyla or archaea, indicating that the bacterial groups lacking this insert have branched off prior to those containing the insert (see interpretive diagram in Supplementary Fig. S1). The presence of this indel in various chlamydiae, V. spinosum and Planctomycetes species, but its absence in different Fibrobacteres-Chlorobi-Bacteroidetes species now provides evidence that the latter groups (FCB) have branched off prior to the Planctomycetes, Verrucomicrobia and Chlamydiales.

**DISCUSSION**

This study was undertaken to clarify the evolutionary relationship of the chlamydiae, Verrucomicrobia and Planctomycetes to each other and other bacterial phyla. The results of these studies provide evidence for a specific relationship between the Verrucomicrobia and...
**Chlamydiales.** In phylogenetic trees based on concatenated protein sequences, all three methods (viz. NJ, ML and MP) supported a specific grouping of the *Verrucomicrobia* and *Chlamydiales* with high bootstrap scores. Further evidence that these two groups shared a common ancestor exclusive of all other bacteria is provided by the signature sequences in the LysRS and RpoB proteins. In these two ubiquitous proteins, conserved inserts of identical lengths were uniquely present in species from these two groups, indicating that they are specifically related. In contrast to the *Verrucomicrobia*, a grouping of the *Planctomycetales* with the chlamydiae-*Verrucomicrobia* clade was statistically supported only in the NJ tree and a weak support for this was also seen in the MP tree. The ML tree did not support a specific affiliation of these groups, where the *Planctomycetales* species grouped with the spirochaetes. We have also not come across any RGC that is uniquely shared by the *Planctomycetales* and the chlamydiae-*Verrucomicrobia* species. Although these studies do not support a specific relationship between the *Planctomycetales* and chlamydiae-*Verrucomicrobia*, species from these groups are indicated to branch in similar positions, as suggested by the species distribution profiles of conserved indels in many other proteins (Supplementary Fig. 1 and other results discussed above). Thus, it is possible that in the future some novel RGCs might be discovered that are uniquely shared by these three groups. However, until such a specific relationship is demonstrated, the placement of *Planctomyces* in a single superphylum with *Verrucomicrobia* and *Chlamydiales* is not warranted (Wagner & Horn, 2006).

All known chlamydiae species are obligate intracellular parasites of the eukaryotic cells (Everett et al., 1999; Kalman et al., 1999; Schachter & Stamm, 1999; Corsaro & Greub, 2006). A specific relationship between them and the *Verrucomicrobia*, as suggested by this work, indicates that *Verrucomicrobia* are the closest free-living relatives of the chlamydiae. In view of this, it is of much interest to determine what other biochemical and physiological proteins are uniquely shared by these two groups (or present in *Verrucomicrobia* but lacking in the chlamydiae), which could provide important insights into the adaptation of the chlamydiae to the parasitic life-style. The signature sequences that are uniquely shared by these groups should also prove useful in the identification of other species that are specifically related to them, some of which may turn out to be even closer relatives of the chlamydiae.

**ACKNOWLEDGEMENTS**

This work was supported by a research grant from the Canadian Institute of Health Research to R.S.G. *BLAST* searches against the preliminary sequence data for *V. spinosum*, *B. marina* and *G. obscuriglobus* were carried out at The Institute for Genomic Research web site at http://www.tigr.org. Sequencing of *V. spinosum* and *G. obscuriglobus* was accomplished with support from NSF and DOE, respectively.

**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.


Griffiths, E., Petrich, A. & Gupta, R. S. (2005). Conserved indels in essential proteins that are distinctive characteristics of *Chlamydiaceae* and provide novel means for their identification. *Microbiology* 151, 2647–2657.


Kalman, S., Mitchell, W., Marathe, R., Lammel, C., Fan, J., Hyman, R. W., Olinger, L., Grimwood, J., Davis, R. W. & Stephens, R. S.


