International Journal of Systematic and Evolutionary Microbiology (2004), **54**, 791–801 DOI 10.1099/ijs.0.02913-0

Evaluation of the phylogenetic position of the planctomycete '*Rhodopirellula baltica*' SH 1 by means of concatenated ribosomal protein sequences, DNA-directed RNA polymerase subunit sequences and whole genome trees

Hanno Teeling,[1] Thierry Lombardot,[1] Margarete Bauer,[1] Wolfgang Ludwig[2] and Frank Oliver Glöckner[1]

[1]Max-Planck-Institute for Marine Microbiology, Celsiusstrasse 1, D-28359 Bremen, Germany

[2]Department of Microbiology, Technical University Munich, D-85350 Freising, Germany

Correspondence
Frank Oliver Glöckner
fog@mpi-bremen.de

In recent years, the planctomycetes have been recognized as a phylum of environmentally important bacteria with habitats ranging from soil and freshwater to marine ecosystems. The planctomycetes form an independent phylum within the bacterial domain, whose exact phylogenetic position remains controversial. With the completion of sequencing of the genome of '*Rhodopirellula baltica*' SH 1, it is now possible to re-evaluate the phylogeny of the planctomycetes based on multiple genes and genome trees in addition to single genes like the 16S rRNA or the elongation factor Tu. Here, evidence is presented based on the concatenated amino acid sequences of ribosomal proteins and DNA-directed RNA polymerase subunits from '*Rhodopirellula baltica*' SH 1 and more than 90 other publicly available genomes that support a relationship of the *Planctomycetes* and the *Chlamydiae*. Affiliation of '*Rhodopirellula baltica*' SH 1 and the *Chlamydiae* was reasonably stable regarding site selection since, during stepwise filtering of less-conserved sites from the alignments, it was only broken when rigorous filtering was applied. In a few cases, '*Rhodopirellula baltica*' SH 1 shifted to a deep branching position adjacent to the *Thermotoga*/*Aquifex* clade. These findings are in agreement with recent publications, but the deep branching position was dependent on site selection and treeing algorithm and thus not stable. A genome tree calculated from normalized BLASTP scores did not confirm a close relationship of '*Rhodopirellula baltica*' SH 1 and the *Chlamydiae*, but also indicated that the *Planctomycetes* do not emerge at the very root of the *Bacteria*. Therefore, these analyses rather contradict a deep branching position of the *Planctomycetes* within the bacterial domain and reaffirm their earlier proposed relatedness to the *Chlamydiae*.

INTRODUCTION

The phylum *Planctomycetes* (Garrity *et al.*, 2002) consists of bacteria whose members share an distinct cell morphology, making them unique within the bacterial domain. Characteristic of the *Planctomycetales* are a polar cell organization and a life cycle with a yeast-like budding mechanism (Schlesner, 1994). The cell walls of planctomycetes are composed of proteins rather than peptidoglycan (König *et al.*, 1984; Liesack *et al.*, 1986; Giovannoni *et al.*, 1987). These cell walls exhibit crateriform structures, small pits that appear as electron-dense circular regions either on the reproductive pole (*Pirellula* species) or on the entire cellular surface (*Planctomyces* species) (Liesack *et al.*, 1986). The most striking morphological feature of planctomycetes, however, is their compartmentalization. The cytoplasm of planctomycetes is divided by an intracytoplasmic membrane into the peripheral, ribosome-free paryphoplasm and the inner, ribosome-containing riboplasm (Lindsay *et al.*, 1997, 2001). The DNA of planctomycetes is highly condensed and forms a nucleoid within the riboplasm, which, in the case of *Gemmata* species, is surrounded by an additional double membrane (Lindsay *et al.*, 2001).

Planctomycetes are widespread and of environmental importance (Fuerst, 1995; Ward-Rainey *et al.*, 1996; Gade *et al.*, 2004). They have been found to be abundant in various

Published online ahead of print on 19 March 2004 as DOI 10.1099/ijs.0.02913-0.

Abbreviations: LBA, long-branch attraction; LGT, lateral gene transfer; LUCA, last universal common ancestor; RBM, reciprocal best match.

Details of the strains used in tree construction are available as supplementary material in IJSEM Online.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

habitats including terrestrial and aquatic habitats differing in salinity (from hypersaline to freshwater), oxygen availability (from the oxic water-column to anoxic sediments), trophic level (from oligotrophic lakes to eutrophic wastewater) and temperature (from cold-water marine snow to hot springs) (Giovannoni *et al.*, 1987; Kerger *et al.*, 1988; DeLong, 1993; Schlesner, 1994; Ward *et al.*, 1995; Vergin *et al.*, 1998; Miskin *et al.*, 1999; Wang *et al.*, 2002). Planctomycetes have even been isolated from the digestive tracts of crustaceans (Fuerst, 1995; Fuerst *et al.*, 1997).

In addition, planctomycetes have interesting metabolic capabilities, e.g. the postulated anammox process, the anaerobic conproportionation of ammonia and nitrite to dinitrogen (Strous *et al.*, 1999; Schmid *et al.*, 2001, Dalsgaard & Thamdrup, 2002).

Despite their distinctive morphology, ubiquitous occurrence and interesting physiology, the phylogeny of planctomycetes still awaits resolution. All studies conducted so far agree on the phylogenetic distinctness of the planctomycetes (Bomar *et al.*, 1988; Ward *et al.*, 2000), but they disagree on the position of the phylum within the tree of life. Early analyses based on 16S rRNA sequences suggested a distant relationship to the *Chlamydiae* (Weisburg *et al.*, 1986; Liesack *et al.*, 1992), whereas such a relationship could not be confirmed in later studies based on 16S/23S rRNA (Ward *et al.*, 2000), *dnaK* (Ward-Rainey *et al.*, 1997) and EF-Tu (Jenkins & Fuerst, 2001). The broad level of sequence divergence within the 5S and 16S rRNA genes of planctomycetes has been interpreted either as an indication that they are rapidly evolving (i.e. contain tachytelic DNA) (Woese, 1987; Bomar *et al.*, 1988; Liesack *et al.*, 1992) or that they represent a very deep-branching phylum (Stackebrandt *et al.*, 1984). In two recent studies based on the slowly evolving positions of the 16S rRNA gene, the *Planctomycetes* have even been described as the deepest branching phylum within the bacterial domain (Brochier & Philippe, 2002) or as branching deeply after the *Thermotoga*/*Aquifex* clade (Di Giulio, 2003).

With the recent completion of the sequencing of the genome of *Pirellula* sp. strain 1 (Glöckner *et al.*, 2003), which will shortly be reclassified as *Rhodopirellula baltica* (Schlesner *et al.*, 2004), and the availability of more than 100 publicly available complete genome sequences, we are now, for the first time, in a position to exploit the wealth of information emerging from entire genomes to reassess the phylogeny of the *Planctomycetes*. In this study, the results of two genomic approaches for phylogenetic tree reconstruction are compared: concatenation of the amino acid sequences of subunits of large information-processing proteins (ribosomal and DNA-directed RNA polymerase subunits) and genome trees based on normalized BLASTP scores (Clarke *et al.*, 2002).

METHODS

Sequences. The amino acid sequences of ribosomal proteins and DNA-directed RNA polymerase subunits were extracted from all bacterial genome sequences that were publicly available on the NCBI website in mid-2003 (ftp://ftp.ncbi.nlm.nih.gov/genomes/Bacteria/). Screening for ribosomal proteins was done with corresponding Pfam profiles (http://www.sanger.ac.uk/Software/Pfam/) using hmmer 2.2g and with BLASTP (ftp://ftp.ncbi.nih.gov/blast/). Remaining subunits were extracted directly from genome sequence annotations. Subunits of the DNA-directed RNA polymerase were searched for with BLASTP or extracted directly from genome sequence annotations.

For the genome tree approach, EMBL-formatted annotated genome sequences were obtained from the EMBL website (ftp://ftp.ebi.ac.uk/pub/databases/embl/genomes) and imported into a local installation of the GenDB annotation system for further analysis (Meyer *et al.*, 2003). The final dataset in the GenDB-MySQL database comprised 85 species accounting for 231 509 ORFs.

Alignments. For analysis of ribosomal proteins, sequences that are known to be prone to lateral gene transfer (LGT), that have paralogues or that were absent in some of the species were excluded from further analysis. The resulting dataset comprised sequences of the following 39 ribosomal proteins from 90 bacterial species: *rpl1–rpl4*, *rpl6*, *rpl7/12*, *rpl9–rpl11*, *rpl13–rpl23*, *rpl27*, *rpl29*, *rpl34*, *rps2–rps9*, *rps11–rps13*, *rps15* and *rps17–rps20*. These were aligned independently using CLUSTAL W 1.83 (with settings gapopen 10 and gapext 0·2) and subsequently concatenated using a custom-made PERL script (9377 aa positions).

For analysis of DNA-directed RNA polymerases, the amino acid sequences of the main subunits *rpoA*, *rpoB* and *rpoC* were extracted from whole genome sequences of 94 bacterial species. The other subunits were left out because *rpoC*1 is restricted to cyanobacteria, *rpoE* is restricted to Gram-positives and *rpoZ* is rather small and seems to be absent from many genomes. The sequences of the main subunits were aligned independently (CLUSTAL W 1.83 with settings gapopen 10 and gapext 0·2) and then concatenated (5277 aa positions).

Columns at which gaps were maximal were omitted from both initial alignments. Afterwards, nine filtered alignments were derived from each initial alignment by successively discarding columns with less than 10, 20, 30, 40, 50, 60, 70, 80 and 90 % sequence conservation.

For analysis with MrBayes (see below), the alignment with 30 % positional conservation filtering was chosen. Species with ambiguously aligned stretches of sequence were removed from the datasets, resulting in final alignments of 82 species for the ribosomal proteins and 92 species for the DNA-directed RNA polymerase subunits.

Phylogenetic analysis. For each alignment, neighbour-joining, parsimony and maximum-likelihood trees were calculated using the programs PROTDIST/NEIGHBOR, PROTPARS and ProML from the PHYLIP 3.6a4 package (http://evolution.genetics.washington.edu/phylip.html) and Pfaat (http://pfaat.sourceforge.net/) (Johnson *et al.*, 2003). The PHYLIP programs were used with default settings and, within Pfaat, neighbour-joining trees were calculated using the BLOSUM62 substitution matrix and global column conservation weighting. Bootstrapping of neighbour-joining and parsimony trees was carried out with 100 replicates. Bootstrapping of the ProML trees was impossible because of the tremendous requirements in memory and computing power. In order to assess the branch support given by a likelihood-based method, trees were calculated for the alignments with 30 % positional conservation filtering using MrBayes version 3 (http://morphbank.ebc.uu.se/mrbayes/) (Huelsenbeck & Ronquist, 2001; Ronquist & Huelsenbeck, 2003). MrBayes uses Bayesian inference estimations to assess phylogeny, is sufficiently fast to allow branch-support evaluation by posterior probabilities and has been shown to be one of the most accurate likelihood programs available (Williams & Moret, 2003). Analysis was carried out using the Jones amino acid substitution model, four chains and an approximated

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

gamma distribution of evolutionary rates with four categories. Visualization of trees was accomplished with ARB (http://www.arb-home.de/).

Genome trees. Genome trees were calculated from normalized BLASTP scores (Clarke *et al.*, 2002). In brief, all 231 509 ORFs were searched against each other using BLASTP. Different substitution matrices were tested (BLOSUM62, PAM70, PAM250). From the results, only the reciprocal BLASTP hits were extracted to avoid paralogous hits. ORFs involved in fewer than a given number of RBMs (reciprocal best matches between genome pairs with an E-value of 10^{-10} or better) were also omitted because they contain too little information (thresholds of 0, 4 and 10 were tested). No filtering of putatively laterally transferred ORFs was applied, because its effect has been proven to be small (Clarke *et al.*, 2002). The remaining data were transferred into a distance matrix as follows. For each ORF in a given query genome, the bit-score of the RBM for this ORF in a given target genome was divided by the self-matching bit-score for the ORF. The mean of these values for a given query–target pair was used as a measure of the overall sequence similarity between the two. Distances were calculated as 1·0 minus the above-mentioned similarity measure. The tree was calculated from the distance matrix with the program FITCH from the PHYLIP 3.6a4 package (model of Fitch–Margoliash, global rearrangements, jumble 100).

RESULTS AND DISCUSSION

In recent years, it has become apparent that the extent of LGT is so great that it must be regarded as one of the major driving forces of evolution. The view that all genetic information in a given lineage traces back to one common ancestor simply does not apply in the world of prokaryotes, where each protein has its own history. In this regard, the tree of life is a complex network of vertical and horizontal inheritance. While LGT is not specifically problematic for the phylogeny of closely related species, it becomes more severe when distantly related organisms are compared. The key question is, whether the extent of LGT is so great that it is impossible to trace the evolutionary relationships between the major lineages. While some scientists believe that the degree of LGT is too great to trace organismal phylogeny on the basis of whole genomes or protein-coding genes (Nesbo *et al.*, 2001), others believe that there is a small set of proteins that forms a robust genetic core of an organism and carries signals of their evolutionary inheritance (Jain *et al.*, 1999; Wolf *et al.*, 2002; Daubin *et al.*, 2003). Part of this core is built by so-called informational proteins. These are proteins of the transcription and translation apparatus, like the subunits of the ribosome and those of the DNA-directed RNA polymerase (Harris *et al.*, 2003). According to the complexity hypothesis (Jain *et al.*, 1999), the physiological interactions and dependencies of these proteins are thought to be much more interwoven than those of operational genes. Therefore, once transferred to another organism via LGT, informational genes are unlikely to be capable of replacing their counterparts. They simply would not fit into the fine-tuned regulation network of their new hosts. Because of their absence of function, they would rapidly accumulate mutations that would first render them inactive and finally cause them to vanish from their hosts' genomes. Despite the fact that the complexity hypothesis is debatable (Nesbo *et al.*, 2001; Daubin *et al.*, 2001), the informational ribosomal and DNA-directed RNA polymerase genes seem to be the best-suited genes, encoding multisubunit proteins whose sequences can be concatenated to infer phylogeny. In addition, it has been demonstrated recently that the extent to which LGT affects typical phylogenetic protein markers might have been overestimated (Daubin *et al.*, 2003). Concatenated ribosomal protein sequences have been applied successfully in a number of phylogenetic studies (Hansmann & Martin, 2000; Wolf *et al.*, 2001; Brochier *et al.*, 2002; Matte-Tailliez *et al.*, 2002; Forterre *et al.*, 2002). They are supposed to have very strong resolving power in evaluating close and intermediate evolutionary distances, i.e. the relationships between species and between major lineages (Wolf *et al.*, 2002).

Besides concatenation of protein sequences, three different methods to infer phylogeny from coding sequences of entire genomes have been developed in recent years. These methods are based on gene content (i.e. presence/absence of genes), gene order and normalized distances between orthologues (Wolf *et al.*, 2002). The gene-content approach was not considered for this study as it is affected by artefacts caused by gene loss. For example, parasitic bacteria with reduced genomes are artificially clustered in gene-content trees (Wolf *et al.*, 2002). Gene-order trees seemed inappropriate because gene order in general is only poorly conserved, which is especially problematic with only one planctomycete genome and no close relative available. Therefore, trees based on normalized BLASTP scores were chosen.

Concatenated ribosomal proteins

Up to a positional conservation filtering of 30 %, all trees calculated from concatenated ribosomal protein sequences successfully resolved the major phyla and, in general, confirmed the currently accepted 16S rRNA-based phylogeny. The *Spirochaetes* and *Chlamydiae* formed a distinct superclade in the likelihood-based (Fig. 1a) and parsimony trees, whereas, in some of the corresponding neighbour-joining trees, these groups formed neighbouring but independent clades (data not shown). The positions of *Chlorobium tepidum* TLS^T, *Deinococcus radiodurans* R1^T and *Thermoanaerobacter tengcongensis* MB4^T were not stable among the trees. *Chlorobium tepidum* TLS^T either affiliated to the *Spirochaetes* or branched before the *Spirochaetes*/*Chlamydiae* clade in the neighbour-joining trees, whereas this species branched before the '*Epsilonproteobacteria*' in the likelihood-based trees and clustered with the '*Epsilonproteobacteria*' in the parsimony trees. The position of *Thermoanaerobacter tengcongensis* MB4^T was dependent on the positional filtering. This species branched next to the *Thermotoga*/*Aquifex* clade with the 10 % positional conservation filter and affiliated to the *Firmicutes* with the 20 and 30 % filters. The position of *Deinococcus radiodurans* R1^T varied considerably, but it affiliated to the cyanobacteria/*Actinobacteria* clade in the majority of trees.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

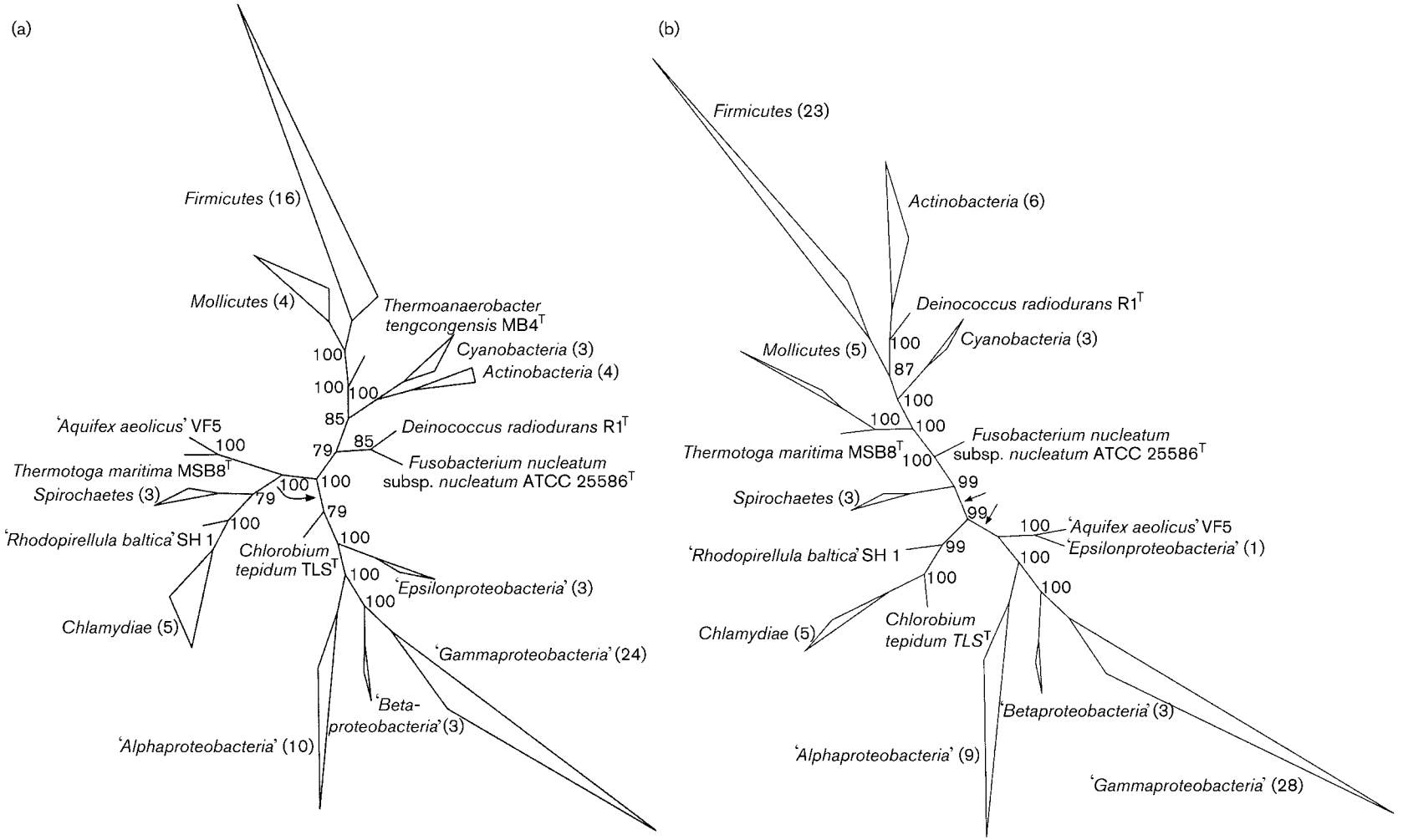

Fig. 1. Unrooted maximum-likelihood trees derived from concatenated protein subunit sequences (30 % positional conservation filtering) with MrBayes using the Jones amino acid substitution model, four chains and an approximated gamma distribution of evolutionary rates with four categories. Numbers in parentheses refer to the numbers of species that make up the respective branches. Details of species used for tree construction are available as supplementary material in IJSEM Online. (a) Consensus (based on 1548 individual trees) derived from concatenated ribosomal protein sequences. The arrow indicates an alternative position of the branch indicated that was present only in the corresponding neighbour-joining tree and that is in better agreement with the currently accepted 16S rRNA-derived phylogeny. (b) Consensus tree (based on 1975 individual trees) derived from concatenated amino acid sequences of DNA-directed RNA polymerases. The arrows indicate alternative positions for '*Rhodopirellula baltica*' SH 1 that were obtained with some of the tested alignments using the parsimony and neighbour-joining methods.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

These exceptions aside, the tree topologies exhibited only little dependence on the treeing algorithms used and all trees consistently placed '*Rhodopirellula baltica*' SH 1 at the base of *Chlamydiae*. Bootstrap support for the node grouping '*Rhodopirellula baltica*' SH 1 with the *Chlamydiae* calculated from the alignment with 30 % positional conservation filtering was 97 (Pfaat), 63 (PROTDIST/NEIGHBOR) and 56 (PROTPARS). The posterior probability in the corresponding MrBayes analysis was 100 (Fig. 1a). The *Planctomycetes–Chlamydiae* relationship was very stable regarding site selection. It was persistent up to a positional conservation filtering of 70 % in the PROTDIST/NEIGHBOR neighbour-joining, 50 % in the Pfaat neighbour-joining, 40 % in the ProML maximum-likelihood and 30 % in the PROTPARS parsimony trees (Table 1). Stricter filtering led to topologies that were often not in agreement with the currently accepted 16S rRNA-based phylogeny, especially regarding the position of the *Mollicutes* and '*Epsilonproteobacteria*'. Likewise, the position of '*Rhodopirellula baltica*' SH 1 became less stable and the association with the *Chlamydiae* was partly lost (Table 1). In three cases, '*Rhodopirellula baltica*' SH 1 shifted to the presumed root of the bacteria, adjacent to the *Thermotoga*/*Aquifex* clade. This position, however, was dependent on the treeing algorithm and site selection. For example, in the parsimony analysis, '*Rhodopirellula baltica*' SH 1 shifted to a deep branching position when a positional conservation filtering of 60 % was applied. With the stricter 70 % filter, however, '*Rhodopirellula baltica*' SH 1 swapped back into the *Spirochaetes*/*Chlamydiae* cluster (Table 1).

Concatenated DNA-directed RNA polymerase subunits

Trees based on concatenated amino acid sequences of DNA-directed RNA polymerase subunits resolved all major lineages known from 16S rRNA trees. The *Spirochaetes* and *Chlamydiae* formed distinct clades in all trees. Up to a positional conservation filtering of 40 %, '*Rhodopirellula baltica*' SH 1 branched together with *Chlorobium tepidum* TLS^T from the chlamydial clade in all likelihood-based trees (Fig. 1b). However, in the majority of parsimony and neighbour-joining trees, '*Rhodopirellula baltica*' SH 1 branched independently either between the *Spirochaetes* and the *Chlamydiae* or between the *Chlamydiae* and the '*Espilonproteobacteria*' (Table 2). Since the position of '*Rhodopirellula baltica*' SH 1 relative to the *Spirochaetes* and *Chlamydiae* was dependent on site selection as well as on the treeing algorithm, a consensus tree would place these lineages within one multifurcating node. Likelihood-based treeing algorithms, however, must be regarded as superior to neighbour-joining and parsimony approaches. Therefore, using not too strict positional conservation filtering, concatenated amino acid sequences of the DNA-directed RNA polymerase support a relationship between the *Planctomycetes* and the *Chlamydiae*, albeit with less strength than concatenated amino acid sequences of ribosomal proteins.

The positions of *Chlorobium tepidum* TLS^T, the cyanobacteria and *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586^T were not stable among the trees. *Chlorobium tepidum* TLS^T affiliated either with '*Rhodopirellula baltica*' SH 1, the *Chlamydiae* or the *Spirochaetes*. *Fusobacterium nucleatum* subsp. *nucleatum* ATCC 25586^T formed an independent lineage in most, but not in all trees and the cyanobacteria had varying positions within the *Actinobacteria*/*Deinococcus*/*Firmicutes* clade. Thus, the overall topology was much more dependent on site selection compared with trees derived from concatenated ribosomal protein sequences. While most trees in general exhibited an overall topology comparable to the 16S rRNA-based phylogeny, there were some noteworthy exceptions. Most notably, '*Aquifex aeolicus*' VF5 did not cluster with *Thermotoga maritima* MSB8^T at the presumed root of the *Bacteria* but, instead, clustered with the '*Epsilonproteobacteria*' and *Deinococcus radiodurans* R1^T affiliated to the *Actinobacteria*.

Table 1. Stability of the *Planctomycetes*/*Chlamydiae* relationship in phylogenetic analysis regarding the selection of sites from concatenated ribosomal protein sequences

Topologies are summarized as: 1, '*Rhodopirellula baltica*' SH 1 branches at the base of the *Chlamydiae*; 2, '*Rhodopirellula baltica*' SH 1 branches near the presumed root of the *Bacteria* adjacent to the *Thermotoga*/*Aquifex* clade; −, other topology. ND, Not determined.

Filter	Amino acid positions	Pfaat	NEIGHBOR	PROTPARS	ProML	MrBayes
10 %	5275	1	1	1	1	ND
20 %	5238	1	1	1	1	ND
30 %	4781	1	1	1	1	1
40 %	4053	1	1	−	1	ND
50 %	3320	1	1	−	−	ND
60 %	2650	−	1	2	1	ND
70 %	2144	2	1	1	1	ND
80 %	1697	2	−	−	−	ND
90 %	1201	−	−	−	−	ND

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

Table 2. Stability of the *Planctomycetes*/*Chlamydiae* relationship in phylogenetic analysis regarding the selection of sites from concatenated sequences of the DNA-directed RNA polymerase

Topologies are summarized as: 1, '*Rhodopirellula baltica*' SH 1 as well as *Chlorobium tepidum* TLS^T branch at the base of the *Chlamydiae*; 2, '*Rhodopirellula baltica*' SH 1 branches independently between the *Chlamydiae* and '*Epsilonproteobacteria*'/*Aquifex* (either with or without and *Chlorobium tepidum* TLS^T); 3, '*Rhodopirellula baltica*' SH 1 branches independently between the *Spirochaetes* and the *Chlamydiae* (either with or without and *Chlorobium tepidum* TLS^T); 4, '*Rhodopirellula baltica*' SH 1 branches near the presumed root of the *Bacteria* adjacent to the *Thermotoga*/*Aquifex* clade; −, other topology (partly not in accordance with the accepted 16S rRNA-derived topology). ND, Not determined.

Filter	Amino acid positions	Pfaat	NEIGHBOR	PROTPARS	ProML	MrBayes
10 %	2550	1	2	2	1	ND
20 %	2529	1	2	2	1	ND
30 %	2322	3	2	1	1	1
40 %	2037	3	2	2	1	ND
50 %	1697	3	−	1	4	ND
60 %	1452	3	−	2	−	ND
70 %	1266	−	−	1	4	ND
80 %	1072	−	−	−	4	ND
90 %	765	−	−	−	−	ND

Furthermore, the *Mollicutes* were clearly separated from the *Firmicutes* and branched more deeply. Such a separation is consistent with trees based on fused 16S and 23S rRNA sequences (*Brochier et al.*, 2002) and the results of different genome tree approaches (Tekaia *et al.*, 1999). A deeper branching position of the *Mollicutes* has been reported before for concatenated DNA-directed RNA polymerase subunits and has been attributed to an accelerated evolutionary rate and thus long-branch attraction (LBA) (Bocchetta *et al.*, 2000). Since this could cause tree distortions, the dataset underlying Fig. 1(b) was reanalysed without the *Mollicutes*. However, this neither changed the position of '*Rhodopirellula baltica*' SH 1 nor altered the overall tree topology (data not shown).

With a positional conservation filtering of 50 % and higher, most of the trees exhibited topologies that were partly inconsistent with the currently accepted 16S rRNA-derived topology. Interestingly, in the ProML maximum-likelihood analysis, '*Aquifex aeolicus*' VF5 shifted adjacent to *Thermotoga maritima* MSB8^T and '*Rhodopirellula baltica*' SH 1 shifted to a position next to the newly formed *Thermotoga*/*Aquifex* clade (Table 2). This position was not found in the corresponding parsimony and neighbour-joining trees. In general, however, the shift of '*Rhodopirellula baltica*' SH 1 towards a deeper branching position with increased filtering of variable positions was more obvious with concatenated amino acid sequences of DNA-directed RNA polymerase subunits than with those of ribosomal proteins.

Genome trees

The genome tree derived from normalized BLASTP scores successfully resolved all major phyla (Fig. 2). Bootstrapping of the genome tree was not possible because of the enormous processing power and time required for its calculation. Thus, branch length and the overall topology were the only measures to assess the reliability of the tree. While separation of the major phyla was good (long branches), their branching pattern was only poorly resolved (very short branches). As in trees based on 16S rRNA analysis, the two thermophiles '*Aquifex aeolicus*' VF5 and *Thermotoga maritima* MSB8^T emerged at the very root of the *Bacteria*. The *Spirochaetes* and *Chlamydiae* formed a well-resolved superclade, while '*Rhodopirellula baltica*' SH 1 emerged as a long independent branch between the *Actinobacteria* and *Cyanobacteria*. The exact position of '*Rhodopirellula baltica*' SH 1 remained ambiguous, however, because the branches of most phyla were too close together to infer their branching pattern reliably. Variation of the BLASTP scoring matrix (BLOSUM62, PAM70, PAM250) and the threshold for RBM filtering of species (0, 4, 10) retained the same overall topology but did not improve the resolution of the tree (data not shown).

Comparison of trees

All trees derived from concatenated protein sequences consistently placed '*Rhodopirellula baltica*' SH 1 near or at the base of the *Chlamydiae* as long as the underlying alignments were not restricted to the most conserved sites. When only highly conserved sites were used for tree construction, '*Rhodopirellula baltica*' SH 1 eventually shifted towards a deeper branching position. This position, however, was not consistent among different treeing algorithms and, furthermore, was highly dependent on which sites were filtered. It is also noteworthy that the overall branch length (and thus resolution) decreased with increased positional filtering and that the bootstrap support

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

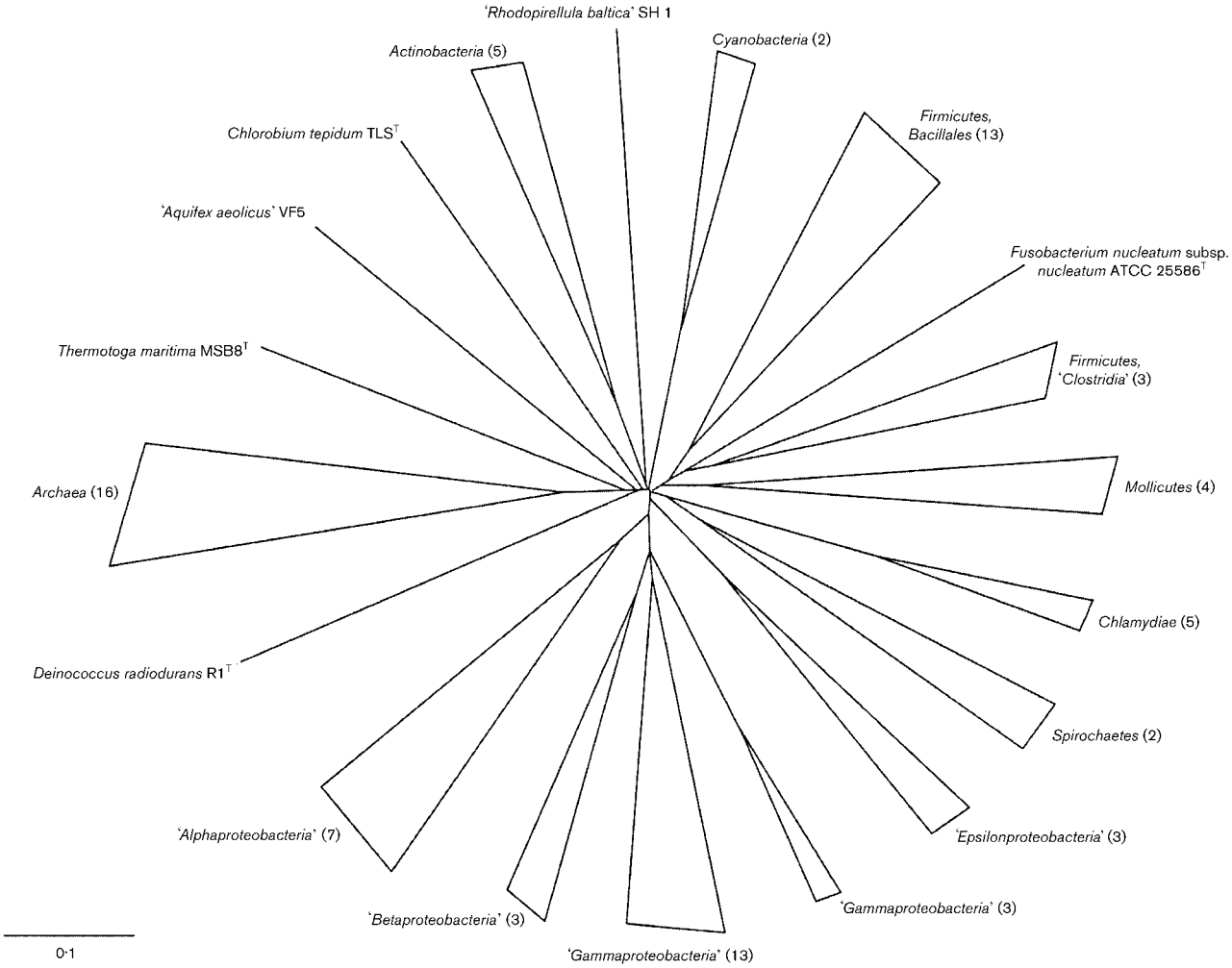

Fig. 2. Genome tree based on normalized BLASTP scores calculated with the Fitch–Margoliash distance matrix method (scoring matrix, BLOSUM62; RBM threshold, $E=10^{-10}$; RBM filter, four species). Numbers in parentheses refer to the number of species that make up the respective branches. Details of species used for tree construction are available as supplementary material in IJSEM Online.

for a deep branching position of '*Rhodopirellula baltica*' SH 1 was low (around 40–50 in the neighbour-joining trees for ribosomal proteins).

The *Chlamydiae* formed a distinct superclade with the *Spirochaetes* in most of the trees derived from concatenated ribosomal protein sequences and in the genome tree, while these two phyla formed neighbouring but independent clades in most trees derived from concatenated DNA-directed RNA polymerase amino acid sequences. The position of the *Chlamydiae* and *Spirochaetes* as neighbours is consistent with the 16S rRNA-based phylogeny, and the two groups forming a distinct superclade is supported by earlier studies on concatenated ribosomal proteins (Wolf *et al.*, 2001, 2002) and by previous genome trees (Clarke *et al.*, 2002).

In contrast to trees inferred from concatenated protein sequences, the genome tree supports neither a close relationship of the planctomycete '*Rhodopirellula baltica*' SH 1 and the *Chlamydiae* nor a deep branching position. However, the exact position of '*Rhodopirellula baltica*' SH 1 is not resolved well in the genome tree. Based on the current dataset, it is impossible to deduce whether '*Rhodopirellula baltica*' SH 1 did not occur within the *Chlamydiae*/*Spirochaetes* superclade because of a lack in resolution or a contradiction to the trees inferred from concatenated protein sequences.

A limitation of all the trees presented here is that, so far, only one complete genome of a planctomycete is publicly available and could be included in tree reconstruction. Clades consisting of a few species are often less stable than those consisting of several species. Thus, inclusion of more than one planctomycete would have been desirable.

The same rationale is valid for the position of *Chlorobium*

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

tepidum TLS^T. This species affiliated with the *Spirochaetes*, '*Rhodopirellula baltica*' SH 1 or the *Chlamydiae* in most of the trees inferred from concatenated protein sequences but not in the genome tree, where its exact position is unclear. Affiliation of *Chlorobium tepidum* TLS^T, the only representative of the *Chlorobi* sequenced so far, with the *Chlamydiae* and *Spirochaetes* is also consistent with the 16S rRNA-based phylogeny (Nelson *et al.*, 2000).

The two thermophiles *Thermotoga maritima* MSB8^T and '*Aquifex aeolicus*' VF5 clustered consistently in trees based on concatenated ribosomal protein sequences and in the genome tree, as they do in the currently accepted 16S rRNA tree. The *Bacteria* are rooted in the genome tree by inclusion of the *Archaea*. Because the two thermophiles emerged at the deepest branching positions within the *Bacteria*, the recently proposed deepest branching position of *Planctomycetes* (Brochier & Philippe, 2002) is not supported by the genome tree.

It is noteworthy that, in the trees inferred from concatenated amino acid sequences of DNA-directed RNA polymerase subunits, the two thermophiles *Thermotoga maritima* MSB8^T and '*Aquifex aeolicus*' VF5 did not cluster when lower positional conservation filtering was applied. Instead, '*Aquifex aeolicus*' VF5 affiliated to the '*Epsilonproteobacteria*'. There is, in fact, a debate going on as to whether, in the currently accepted 16S rRNA-based phylogenetic tree, the thermophiles were placed at the root of the *Bacteria* erroneously due to LBA and whether the last common ancestor of the *Bacteria* was a thermophile at all (Daubin *et al.*, 2001; Gribaldo & Philippe, 2002). It has been proposed previously that '*Aquifex aeolicus*' VF5 is closely related to *Proteobacteria* (Philippe & Laurent, 1998), which is indicated by gene-content trees, for example (Wolf *et al.*, 2002).

The recently proposed *Actinobacteria*/*Cyanobacteria*/*Deinococcus* superclade (Wolf *et al.*, 2001, 2002) was found in most of the trees based on concatenated protein sequences, but was not resolved in the genome tree. It is likely that the resolution power of the genome tree method is blurred due to horizontally transferred genes. The inclusion of the *Archaea* might also have had a limiting effect on the resolution within the *Bacteria*, because genes that were laterally transferred between the two domains minimize the distances within the *Bacteria*.

Additional support for a relationship of the *Planctomycetes* and *Chlamydiae*

A relationship of the *Planctomycetes* and *Chlamydiae* is further supported, albeit weakly, by indels. Indels are conserved insertions and deletions in key proteins that are assumed to be phylum-specific and thus suited for phylogeny. A system based on 18 indels was developed by Gupta and colleagues (Gupta, 2001; Gupta & Griffiths, 2002). According to this system, inserts in the termination factor *rho* and the alanyl-tRNA synthetase (*alaS*) are supposed to be diagnostic for species that arose after the branching of the *Spirochaetes* and the *Chlamydiae*, respectively. While '*Rhodopirellula baltica*' SH 1 carries the first insert, it lacks the latter, which is consistent with a branching between the *Spirochaetes* and *Chlamydiae*. However, like all phylogenetic methods, the indel method has its limitations (Philippe & Laurent, 1998; Gribaldo & Philippe, 2002). These are especially obvious in the case of '*Rhodopirellula baltica*' SH 1, where some of the markers are absent (*ftsZ*, *hsp*90, *lon* protease, inorganic pyrophosphatase), have paralogues (*hemL*, *dnaK*) or are fused with other genes (*secF*).

Further support for an affiliation of the *Planctomycetes* and the *Chlamydiae*/*Spirochaetes* clade comes from trees derived from concatenated sequences of subunits of the well-conserved F_1F_0-ATPase operon (data not shown). In these trees, '*Rhodopirellula baltica*' SH 1 consistently clustered with the only spirochaete in the dataset, *Leptospira interrogans* serovar lai 56601. Chlamydiae do not have an F_1F_0-type ATPase and consequently were absent from the dataset. However, as mentioned above, the *Spirochaetes* and *Chlamydiae* are assumed to be close relatives and thus clustering of '*Rhodopirellula baltica*' SH 1 with the *Spirochaetes* weakly supports an overall affiliation of the *Planctomycetes* with the *Spirochaetes*/*Chlamydiae* superclade. Preliminary phylogenetic analysis of the RecA protein also indicated a close relationship of '*Rhodopirellula baltica*' SH 1 and the *Chlamydiae* (data not shown).

Aside from the results of phylogenetic analysis, it seems at first rather surprising that the *Chlamydiae* and *Planctomycetes* should have evolved from a common ancestor. Chlamydiae are small, intracellular energy parasites with reduced genomes, while planctomycetes are free-living bacteria with genomes that are among the largest bacterial genomes known. There are, however, some noteworthy analogies between the two groups, such as proteinaceous cell walls that are cross-linked via disulphide bonds. Planctomycetes do have these in general (König *et al.*, 1984; Liesack *et al.*, 1986; Giovannoni *et al.*, 1987) and chlamydiae do during their elementary body state (Hatch, 1996). The existence of all genes required for peptidoglycan biosynthesis in some chlamydiae (Stephens *et al.*, 1998; Ghuysen & Goffin, 1999) and of some of these genes in '*Rhodopirellula baltica*' SH 1 (Glöckner *et al.*, 2003) indicate that both groups once possessed peptidoglycan and that their proteinaceous cell walls are secondary adaptations. Chlamydiae and planctomycetes not only exhibit complex cell cycles but also lack *ftsZ*, indicating an unknown mode of cell division (Brown & Rockey, 2000; Glöckner *et al.*, 2003). Related to their cell division might also be the fact that, of all genomes sequenced so far, only chlamydiae and '*Rhodopirellula baltica*' SH 1 harbour two copies of the gene *dnaA*. Moreover, these genes seem to be distantly related (Glöckner *et al.*, 2003). In addition, the genomes of all six sequenced chlamydiae (Karunakaran *et al.*, 2003) as well as that of '*Rhodopirellula baltica*' SH 1 harbour three copies of

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

groEL-like genes. Furthermore, chlamydiae (Hatch, 1996) and planctomycetes both have highly condensed DNA that is visible in electron micrographs as nucleoids. The 16S rRNA genes of both groups share signature positions that are not present in other bacteria (Fuerst, 1995) and, finally, the ribosomal *spc* operon in all chlamydiae sequenced to date, as well as in '*Rhodopirellula baltica*' SH 1, is devoid of the ribosomal protein L30, as in some other bacteria, e.g. *Synechococcus* sp. PCC 6301 (Sugita *et al.*, 1997).

Relationship to the *Chlamydiae* versus deep branching position

In two aforementioned studies, the slowly evolving positions of the 16S rRNA have been used to infer the phylogeny of the deepest branching species within the bacterial domain. The rationale for this approach is that the phylogenetic signal of very ancient relationships is retained exclusively in the slowly evolving positions but might be obscured by the faster evolving ones if these are not filtered. Brochier & Philippe (2002) showed that the *Planctomycetes* shift to the very root of the *Bacteria* when only 751 slowly evolving positions are used for tree reconstruction and concluded that the last universal common ancestor (LUCA) might not have been a thermophile. In a re-evaluation of their results, Di Giulio (2003) demonstrated that a different selection of slowly evolving sites re-establishes the deepest branching position of the thermophiles (*Thermotoga maritima* MSB8^T, '*Aquifex aeolicus*' VF5) and places the *Planctomycetes* at a deep branching position after the thermophiles (this analysis is, unfortunately, devoid of *Chlamydiae*). The nature of the LUCA is beyond the scope of this study. Regarding the position of the *Planctomycetes*, however, our data also exhibit an undeniable, albeit inconsistent, tendency to place the *Planctomycetes* at a deeper branching position within the *Bacteria* when only highly conserved positions are used for tree reconstruction. Both of the aforementioned studies used parsimony analysis to select the slowly evolving positions, while our analysis is based on a general filtering of variable positions. We therefore cannot exclude the possibility that a different mode of site selection from our alignments would have led to a more stable deep-branching position of the *Planctomycetes*. However, one might ask whether filtering of the majority of unambiguously aligned positions from an alignment really reveals an otherwise obscured phylogenetic signal or if it rather introduces artefacts. In the end, the association of '*Rhodopirellula baltica*' SH 1 with the *Chlamydiae* in our phylogenetic analysis was quite consistent, especially for the very large alignment of ribosomal protein sequences. Ignoring this signal in favour of a weakly supported deep branching position would be hard to justify. It is also noteworthy that, in numerous trees that were calculated from alignments with very strict positional conservation filtering, other species (e.g. *Campylobacter jejuni* subsp. *jejuni* NCTC 11168) also often shifted to a deep branching position. Finally, a deep branching position of the *Planctomycetes* fails to explain the numerous analogies that exist between the *Planctomycetes* and the *Chlamydiae*. Therefore, our analyses support a close affiliation of the *Planctomycetes* with the *Chlamydiae* rather than a deep branching position of the *Planctomycetes*.

Conclusions

The fact that different markers like concatenated protein sequences of ribosomal and DNA-directed RNA polymerase subunits reveal comparable overall topologies encourages the view that, despite LGT, prokaryote genomes retain a phylogenetic signal from which relationships can be reliably inferred. Phylogenetic analysis of concatenated amino acid sequences of ribosomal and DNA-directed RNA polymerase subunits concordantly indicates a close relationship of the *Planctomycetes* and *Chlamydiae*. It seems unlikely that both groups of proteins are affected at the same time by LGT from chlamydiae that mimics an otherwise non-existent relationship. This case, however, can of course not be excluded with certainty. Also, a false affiliation of the two groups due to LBA is possible. These scenarios, however, seem unlikely, since the relationship of the *Planctomycetes* and *Chlamydiae* is further supported, albeit with varying strength, by phylogenetic analysis of RecA, indels and some noteworthy analogies between the two groups. Phylogenetic analysis of concatenated ATPase subunits weakly supported an affiliation of *Planctomycetes* and *Spirochaetes* and thus the *Chlamydiae*/*Spirochaetes* superclade.

With respect to resolution power, concatenation of protein sequences of ribosomal and DNA-directed RNA polymerase subunits did provide better resolution for distant relationships than genome trees. Whether, as suggested by some authors (Wolf *et al.*, 2001), concatenated sequences of ribosomal proteins are superior to 16S rRNA-based phylogeny in assessing the overall topology of the bacterial tree of life cannot be deduced from our data. As with all protein-based phylogenies, concatenation of protein sequences has to face the problems of LGT and paralogy. In addition, site selection has a major impact on the weakly supported branches of the inferred trees, which especially affects the position of the *Chlamydiae* (Hansmann & Martin, 2000). However, the *Planctomycetes*–*Chlamydiae* relationship in our trees based on concatenated protein sequences was quite stable regarding site selection. In addition, trees based on concatenated sequences of ribosomal proteins from different workgroups show only slight differences in their branching patterns and are remarkably similar (Wolf *et al.*, 2002). Furthermore, 16S rRNA-based phylogeny also has to face the problem of paralogy, since most bacterial genomes harbour more than one set of rRNA genes. In addition, there are cases where the 16S rRNA sequences within one organism can vary considerably, and even LGT of the 16S rRNA gene has been described (Yap *et al.*, 1999). Moreover, the information content of the 16S rRNA gene is limited. The overall topology of the bacterial tree of life is beyond the scope of this study. However, regarding bootstrap values and branch length (data not shown), concatenated ribosomal proteins

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

and DNA-directed RNA polymerase subunits do seem to have sufficient resolving power to infer the phylogeny of the *Planctomycetes*.

It will be interesting to see whether the phylogenetic relationship of the *Planctomycetes* with the *Chlamydiae* that is indicated by most of our data will hold true as more planctomycete genomes (e.g. *Gemmata obscuriglobus* UQM 2246^T, which is currently sequenced by The Institute for Genomic Research) or those of the recently discovered environmental *Parachlamydia* (http://www.microbial-ecology.net/edge.html) become available to the public. We hope that an in-depth analysis of future planctomycete genomes will help to provide further insights into their phylogenetic position.

ACKNOWLEDGEMENTS

We would like to acknowledge Tim Frana for screening all published bacterial genomes for ribosomal proteins. Major funding of this study was provided by the German Federal Ministry of Education and Research and the Max Planck Society.

REFERENCES

Bocchetta, M., Gribaldo, S., Sanangelantoni, A. & Cammarano, P. (2000). Phylogenetic depth of the bacterial genera *Aquifex* and *Thermotoga* inferred from analysis of ribosomal protein, elongation factor, and RNA polymerase subunit sequences. *J Mol Evol* **50**, 366–380.

Bomar, D., Giovannoni, S. & Stackebrandt, E. (1988). A unique type of eubacterial 5S rRNA in members of the order *Planctomycetales*. *J Mol Evol* **27**, 121–125.

Brochier, C. & Philippe, H. (2002). Phylogeny: a non-hyperthermophilic ancestor for bacteria. *Nature* **417**, 244.

Brochier, C., Bapteste, E., Moreira, D. & Philippe, H. (2002). Eubacterial phylogeny based on translational apparatus proteins. *Trends Genet* **18**, 1–5.

Brown, W. J. & Rockey, D. D. (2000). Identification of an antigen localized to an apparent septum within dividing chlamydiae. *Infect Immun* **68**, 708–715.

Clarke, G. D. P., Beiko, R. G., Ragan, M. A. & Charlebois, R. L. (2002). Inferring genome trees by using a filter to eliminate phylogenetically discordant sequences and a distance matrix based on mean normalized BLASTP scores. *J Bacteriol* **184**, 2072–2080.

Dalsgaard, T. & Thamdrup, B. (2002). Factors controlling anaerobic ammonium oxidation with nitrite in marine sediments. *Appl Environ Microbiol* **68**, 3802–3808.

Daubin, V., Gouy, M. & Perrière, G. (2001). Bacterial molecular phylogeny using supertree approach. *Genome Inform Ser Workshop Genome Inform* **12**, 155–164.

Daubin, V., Moran, N. A. & Ochman, H. (2003). Phylogenetics and the cohesion of bacterial genomes. *Science* **301**, 829–832.

DeLong, E. F. (1993). Phylogenetic diversity of aggregate-attached vs. free-living marine bacterial assemblages. *Limnol Oceanogr* **38**, 924–934.

Di Giulio, M. (2003). The ancestor of the *Bacteria* domain was a hyperthermophile. *J Theor Biol* **224**, 277–283.

Forterre, P., Brochier, C. & Philippe, H. (2002). Evolution of the *Archaea*. *Theor Pop Biol* **61**, 409–422.

Fuerst, J. A. (1995). The planctomycetes: emerging models for microbial ecology, evolution and cell biology. *Microbiology* **141**, 1493–1506.

Fuerst, J. A., Gwilliam, H. G., Lindsay, M., Lichanska, A., Belcher, C., Vickers, J. E. & Hugenholtz, P. (1997). Isolation and molecular identification of planctomycete bacteria from postlarvae of the giant tiger prawn, *Penaeus monodon*. *Appl Environ Microbiol* **63**, 254–262.

Gade, D., Schlesner, H., Glöckner, F. O., Amann, R., Pfeiffer, S. & Thomm, M. (2004). Identification of planctomycetes with order-, genus- and strain-specific 16S rRNA-targeted probes. *Microb Ecol* (in press). Published online ahead of print 4 March 2004 as DOI 10.1007/s00248-003-1016-9

Garrity, G. M., Johnson, K. L., Bell, J. & Searles, D. B. (2002). Taxonomic outline of the prokaryotes. In *Bergey's Manual of Systematic Bacteriology*, 2nd edn, release 3.0. July 2002. http://dx.doi.org/10.1007/bergeysoutline200210

Ghuysen, J. M. & Goffin, C. (1999). Lack of cell wall peptidoglycan versus penicillin sensitivity: new insights into the chlamydial anomaly. *Antimicrob Agents Chemother* **43**, 2339–2344.

Giovannoni, S. J., Godchaux, W., III, Schabtach, E. & Castenholz, R. W. (1987). Cell wall and lipid composition of *Isosphaera pallida*, a budding eubacterium from hot springs. *J Bacteriol* **169**, 2702–2707.

Glöckner, F. O., Kube, M., Bauer, M. & 11 other authors (2003). Complete genome sequence of the marine planctomycete *Pirellula* sp. strain 1. *Proc Natl Acad Sci U S A* **100**, 8298–8303.

Gribaldo, S. & Philippe, H. (2002). Ancient phylogenetic relationships. *Theor Pop Biol* **61**, 391–408.

Gupta, R. S. (2001). The branching order and phylogenetic placement of species from completed bacterial genomes, based on conserved indels found in various proteins. *Int Microbiol* **4**, 187–202.

Gupta, R. S. & Griffiths, E. (2002). Critical issues in bacterial phylogeny. *Theor Pop Biol* **61**, 423–434.

Hansmann, S. & Martin, W. (2000). Phylogeny of 33 ribosomal and six other proteins encoded in an ancient gene cluster that is conserved across prokaryotic genomes: influence of excluding poorly alignable sites from analysis. *Int J Syst Evol Microbiol* **50**, 1655–1663.

Harris, J. K., Kelley, S. T., Spiegelman, G. B. & Pace, N. R. (2003). The genetic core of the universal ancestor. *Genome Res* **13**, 407–412.

Hatch, T. P. (1996). Disulfide cross-linked envelope proteins: the functional equivalent of peptidoglycan in chlamydiae? *J Bacteriol* **178**, 1–5.

Huelsenbeck, J. P. & Ronquist, F. (2001). MrBayes: Bayesian inference of phylogenetic trees. *Bioinformatics* **17**, 754–755.

Jain, R., Rivera, M. C. & Lake, J. A. (1999). Horizontal gene transfer among genomes: the complexity hypothesis. *Proc Natl Acad Sci U S A* **96**, 3801–3806.

Jenkins, C. & Fuerst, J. A. (2001). Phylogenetic analysis of evolutionary relationships of the planctomycete division of the domain *Bacteria* based on amino acid sequences of elongation factor Tu. *J Mol Evol* **52**, 405–418.

Johnson, J. M., Mason, K., Moallemi, C., Xi, H., Somaroo, S. & Huang, E. S. (2003). Protein family annotation in a multiple alignment viewer. *Bioinformatics* **19**, 544–545.

Karunakaran, K. P., Noguchi, Y., Read, T. D., Cherkasov, A., Kwee, J., Shen, C., Nelson, C. C. & Brunham, R. C. (2003). Molecular analysis of the multiple GroEL proteins of *Chlamydiae*. *J Bacteriol* **185**, 1958–1966.

Kerger, D., Mancuso, A., Nichols, P. D., White, D. C., Langworthy, T., Sittig, M., Schlesner, H. & Hirsch, P. (1988). The budding bacteria, *Pirellula* and *Planctomyces*, with atypical 16S rRNA and absence of peptidoglycan, show eubacterial phospholipids und uniquely high

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15

proportions of long chain beta-hydroxy fatty acids in the lipopolysaccharide lipid A. *Arch Microbiol* **149**, 255–260.

König, E., Schlesner, H. & Hirsch, P. (1984). Cell wall studies on budding bacteria of the *Planctomyces*/*Pasteuria* group and on a *Prosthecomicrobium* sp. *Arch Microbiol* **138**, 200–205.

Liesack, W., König, H., Schlesner, H. & Hirsch, P. (1986). Chemical composition of the peptidoglycan-free cell envelopes of budding bacteria of the *Pirellula*/*Planctomyces* group. *Arch Microbiol* **145**, 361–366.

Liesack, W., Söller, R., Steward, T., Haas, H., Giovannoni, S. & Stackebrandt, E. (1992). The influence of tachytelically (rapidly) evolving sequences on the topology of phylogenetic trees – intrafamily relationships and the phylogenetic position of *Planctomycetaceae* as revealed by comparative analysis of 16S ribosomal RNA sequences. *Syst Appl Microbiol* **15**, 357–362.

Lindsay, M. R., Webb, R. I. & Fuerst, J. A. (1997). Pirellulosomes: a new type of membrane-bounded cell compartment in planctomycete bacteria of the genus *Pirellula*. *Microbiology* **143**, 739–748.

Lindsay, M. R., Webb, R. I., Strous, M., Jetten, M. S., Butler, M. K., Forde, R. J. & Fuerst, J. A. (2001). Cell compartmentalisation in planctomycetes: novel types of structural organisation for the bacterial cell. *Arch Microbiol* **175**, 413–429.

Matte-Tailliez, O., Brochier, C., Forterre, P. & Philippe, H. (2002). Archaeal phylogeny based on ribosomal proteins. *Mol Biol Evol* **19**, 631–639.

Meyer, F., Goesmann, A., McHardy, A. C. & 8 other authors (2003). GenDB – an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res* **31**, 2187–2195.

Miskin, I. P., Farrimond, P. & Head, I. M. (1999). Identification of novel bacterial lineages as active members of microbial populations in a freshwater sediment using a rapid RNA extraction procedure and RT-PCR. *Microbiology* **145**, 1977–1987.

Nelson, K. E., Paulsen, I. T., Heidelberg, J. F. & Fraser, C. M. (2000). Status of genome projects for nonpathogenic bacteria and archaea. *Nat Biotechnol* **18**, 1049–1054.

Nesbo, C. L., Boucher, Y. & Doolittle, W. F. (2001). Defining the core of nontransferable prokaryotic genes: the euryarchaeal core. *J Mol Evol* **53**, 340–350.

Philippe, H. & Laurent, J. (1998). How good are deep phylogenetic trees? *Curr Opin Genet Dev* **8**, 616–623.

Ronquist, F. & Huelsenbeck, J. P. (2003). MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**, 1572–1574.

Schlesner, H. (1994). The development of media suitable for the microorganisms morphologically resembling *Planctomyces* spp., *Pirellula* spp. and other *Planctomycetales* from various aquatic habitats using dilute media. *Syst Appl Microbiol* **17**, 135–145.

Schlesner, H., Rendsmann, C., Tindall, B. J., Gade, D., Rabus, R., Pfeiffer, S. & Hirsch, P. (2004). Taxonomic heterogeneity within the *Planctomycetales* as derived by DNA–DNA hybridization, description of *Rhodopirellula baltica* gen. nov., sp. nov., transfer of *Pirellula marina* to the genus *Blastopirellula* gen. nov. as *Blastopirellula marina* comb. nov., and emended description of the genus *Pirellula*. *Int J Syst Evol Microbiol* **54** (in press).

Schmid, M., Schmitz-Esser, S., Jetten, M. & Wagner, W. (2001). 16S-23S rDNA intergenic spacer and 23S rDNA of anaerobic ammonium-oxidizing bacteria: implications for phylogeny and *in situ* detection. *Environ Microbiol* **3**, 450–459.

Stackebrandt, E., Ludwig, W., Schubert, W., Klink, F., Schlesner, H., Roggentin, T. & Hirsch, P. (1984). Molecular genetic evidence for early evolutionary origin of budding peptidoglycan-less eubacteria. *Nature* **307**, 735–737.

Stephens, R. S., Kalman, S., Lammel, C. & 9 other authors (1998). Genome sequence of an obligate intracellular pathogen of humans: *Chlamydia trachomatis*. *Science* **282**, 754–759.

Strous, M., Fuerst J. A., Kramer, E. H. M., Logemann, S., Muyzer, G., van de pas-Schoonen, K. T., Webb, R., Kuenen, J. G. & Jetten, M. S. M. (1999). Missing lithotroph identified as new planctomycete. *Nature* **400**, 446–449.

Sugita, M., Sugishita, H., Fujishiro, T., Tsuboi, M., Sugita, C., Endo, T. & Sugiura, M. (1997). Organization of a large gene cluster encoding ribosomal proteins in the cyanobacterium *Synechococcus* sp. strain PCC 6301: comparison of gene clusters among cyanobacteria, eubacteria and chloroplast genomes. *Gene* **195**, 73–79.

Tekaia, F., Lazcano, A. & Dujon, B. (1999). The genomic tree as revealed from whole proteome comparisons. *Genome Res* **9**, 550–557.

Vergin, K. L., Urbach, E., Stein, J. L., DeLong, E. F., Lanoil, B. D. & Giovannoni, S. J. (1998). Screening of a fosmid library of marine environmental genomic DNA fragments reveals four clones related to members of the order *Planctomycetales*. *Appl Environ Microbiol* **64**, 3075–3078.

Wang, J., Jenkins, C., Webb, R. I. & Fuerst, J. A. (2002). Isolation of *Gemmata*-like and *Isosphaera*-like bacteria from soil and freshwater. *Appl Environ Microbiol* **68**, 417–422.

Ward, N., Rainey, F. A., Stackebrandt, E. & Schlesner, H. (1995). Unraveling the extent of diversity within the order *Planctomycetales*. *Appl Environ Microbiol* **61**, 2270–2275.

Ward, N. L., Rainey, F. A., Hedlund, B. P., Staley, J. T., Ludwig, W. & Stackebrandt, E. (2000). Comparative phylogenetic analyses of members of the order *Planctomycetales* and the division *Verrucomicrobia*: 23S rRNA gene sequence analysis supports the 16S rRNA gene sequence-derived phylogeny. *Int J Syst Evol Microbiol* **50**, 1965–1972.

Ward-Rainey, N., Rainey, F. A., Wellington, E. M. H. & Stackebrandt, E. (1996). Physical map of the genome of *Planctomyces limnophilus*, a representative of the phylogenetically distinct planctomycete lineage. *J Bacteriol* **178**, 1908–1913.

Ward-Rainey, N., Rainey, F. A. & Stackebrandt, E. (1997). The presence of a *dnaK* (HSP70) multigene family in members of the orders *Planctomycetales* and *Verrucomicrobiales*. *J Bacteriol* **179**, 6360–6366.

Weisburg, W. G., Hatch, T. P. & Woese, C. R. (1986). Eubacterial origin of chlamydiae. *J Bacteriol* **167**, 570–574.

Williams, T. L. & Moret, B. M. E. (2003). An investigation of phylogenetic likelihood methods. In *Proceedings of the Third IEEE Symposium on Bioinformatics and Bioengineering (BIBE '03)*, pp. 79–86. Los Alamitos, CA: IEEE Press. http://www.computer.org/proceedings/bibe/1907/1907toc.htm

Woese, C. R. (1987). Bacterial evolution. *Microbiol Rev* **51**, 221–271.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V., Tatusov, R. L. & Koonin, E. V. (2001). Genome trees constructed using five different approaches suggest new major bacterial clades. *BMC Evol Biol* **1**, 8.

Wolf, Y. I., Rogozin, I. B., Grishin, N. V. & Koonin, E. V. (2002). Genome trees and the tree of life. *Trends Genet* **18**, 472–479.

Yap, W. H., Zhang, Z. & Wang, Y. (1999). Distinct types of rRNA operons exist in the genome of the actinomycete *Thermomonospora chromogena* and evidence for horizontal transfer of an entire rRNA operon. *J Bacteriol* **181**, 5201–5209.

Downloaded from www.sgmjournals.org by
IP: 93.91.26.109
On: Wed, 05 Aug 2015 04:18:15