Identification of Proteins Likely To Be Involved in Morphogenesis, Cell Division, and Signal Transduction in Planctomycetes by Comparative Genomics

Christian Jogler, a b Jost Waldmann, c Xiaolu Huanga, Mareike Jogler, b Frank Oliver Glöckner, c Thorsten Mascher, d and Roberto Kolter a

Harvard Medical School, Boston, Massachusetts, USA a; DSMZ, Braunschweig, Germany b; MPI for Marine Microbiology, and Jacobs University, Bremen, Germany c; and Ludwig Maximilians University, Munich, Germany d

Members of the Planctomycetes clade share many unusual features for bacteria. Their cytoplasm contains membrane-bound compartments, they lack peptidoglycan and PtsZ, they divide by polar budding, and they are capable of endocytosis. Planctomycete genomes have remained enigmatic, generally being quite large (up to 9 Mb), and on average, 55% of their predicted proteins are of unknown function. Importantly, proteins related to the unusual traits of Planctomycetes remain largely unknown. Thus, we embarked on bioinformatic analyses of these genomes in an effort to predict proteins that are likely to be involved in compartmentalization, cell division, and signal transduction. We used three complementary strategies. First, we defined the Planctomycetes core genome and subtracted genes of well-studied model organisms. Second, we analyzed the gene content and synteny of morphogenesis and cell division genes and combined both methods using a “guilt-by-association” approach. Third, we identified signal transduction systems as well as sigma factors. These analyses provide a manageable list of candidate genes for future genetic studies and provide evidence for complex signaling in the Planctomycetes akin to that observed for bacteria with complex life-styles, such as Myxococcus xanthus.

Received 24 July 2012 Accepted 14 September 2012
Published ahead of print 21 September 2012
Address correspondence to Christian Jogler, christian@jogler.de.
Supplemental material for this article may be found at http://jb.asm.org/.
Copyright © 2012, American Society for Microbiology. All Rights Reserved.
doi:10.1128/JB.01325-12
TABLE 1. Selected features of sequenced planctomycetal genomes except for anammox bacteria*  

<table>
<thead>
<tr>
<th>Organism</th>
<th>Chromosome size (Mb)</th>
<th>Sequence type</th>
<th>GenBank accession no.</th>
<th>Plasmid size (kb)</th>
<th>No. of genes</th>
<th>% G+C content</th>
<th>% of PCGs with function</th>
<th>No. of tRNAs</th>
<th>No. of rRNAs</th>
<th>Reference or sources</th>
</tr>
</thead>
<tbody>
<tr>
<td>Rhodopirellula baltica SH1</td>
<td>7.1</td>
<td>Complete</td>
<td>NC_003027.1</td>
<td>7,403</td>
<td>3,725</td>
<td>55.40</td>
<td>34.22</td>
<td>56</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Planctomycetes limnophilus</td>
<td>5.4</td>
<td>Complete</td>
<td>NC_014148.1</td>
<td>4,313 (+60)</td>
<td>3,958 (57.0)</td>
<td>55.70</td>
<td>34.22</td>
<td>56</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Pirellula staleyi</td>
<td>6.2</td>
<td>Complete</td>
<td>NC_013720.1</td>
<td>4,825</td>
<td>5,177</td>
<td>57.50</td>
<td>34.22</td>
<td>56</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Isosphaera pallida</td>
<td>5.5</td>
<td>Complete</td>
<td>NC_014692.1</td>
<td>2,671 (+32)</td>
<td>2,609 (57.0)</td>
<td>53.70</td>
<td>34.22</td>
<td>56</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Planctomycetes brasiliensis</td>
<td>6.0</td>
<td>Draft</td>
<td>NZ_AEC00000000</td>
<td>4,863</td>
<td>5,769</td>
<td>62.00</td>
<td>34.22</td>
<td>56</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Gemmata obscuriglobus</td>
<td>9.1</td>
<td>Draft</td>
<td>NZ_AEC00000000</td>
<td>2,671 (+32)</td>
<td>2,609 (57.0)</td>
<td>53.70</td>
<td>34.22</td>
<td>56</td>
<td>4</td>
<td>6</td>
</tr>
<tr>
<td>Planctomycetes maris</td>
<td>7.8</td>
<td>Draft</td>
<td>NZ_AEC00000000</td>
<td>6,530</td>
<td>6,480</td>
<td>30.50</td>
<td>40.66</td>
<td>48</td>
<td>1</td>
<td>6</td>
</tr>
<tr>
<td>Blastopirellula marina</td>
<td>6.7</td>
<td>Draft</td>
<td>NZ_AAN00000000</td>
<td>6,079</td>
<td>6,025</td>
<td>57.00</td>
<td>39.64</td>
<td>48</td>
<td>1</td>
<td>6</td>
</tr>
</tbody>
</table>

* PCG, protein-coding genes; ND, not determined; CDS, coding sequences. Numbers in parentheses provide data for inclusion of plasmids.

** Downloaded from TrEMBL.

MATERIALS AND METHODS

Phylogenetic analyses of Planctomycetes. Aligned 16S rRNA gene sequences were derived from the SILVA database (release 108 [http://www.arb-silva.de/download/arb-files/) (47), and the alignment was correctly manually. Phylogenetic trees were calculated by employing the ARB software package (31). The RAXML module (rate distribution model GTRGAMMA; rapid bootstrap analysis algorithm) performed the maximum likelihood analysis, while neighbor-joining trees were calculated with the ARB Neighbor Joining tool using Felsenstein correction. Boot-strap values were determined by using 1,000-fold resampling (31). GenBank accession numbers of the sequences used are listed in Table S1 in the supplemental material.

Distilling the planctomycetal core genome. First, an NCBI BLAST database ("core contributors") containing all available protein sequences of Planctomycetes was generated (Table 1), preserving their taxonomic information. Second, an all-against-all BLASTP search was used to extract all homologues within the data set. To exclude matches to the originating genome, the mask option (negative_gilist) of BLAST was used. If a protein matched the same database entry multiple times, only the best BLAST match was considered. Hits with a query coverage higher than 60% and below the E value cutoff of 1e−5 were taken into account. Additionally, the number of species in the associated list of BLAST matches was counted for each query protein. Queries not matching all taxa in the database, except the originating one, were discarded. Subsequently, sequences of queries fulfilling the selection criteria were compiled into a list, and all genes among the list linked by reciprocal BLAST hits were clustered. If a BLAST match did not link back to the query, the corresponding sequence was rejected. Consequently, each cluster represents a group of homologous genes, with a high probability of sharing the same function.

Trimming the planctomycetal core genome using in silico subtractions. We compiled a set of genes from well-studied model organisms (Escherichia coli strain K-12 substrain MG1655 [GenBank accession number U00096]) and Bacillus subtilis subsp. subtilis strain 168 (accession number AL0910921)), and a second BLAST database ("blacklist") was generated out of those. All genes of the planctomycete core genome were compared against this blacklist database by using BLAST. Genes reaching the E value cutoff of 1e−5 and a minimum coverage of 65% led to the deletion of the entire cluster from the planctomycetal core genome set. Corresponding Perl scripts are available upon request. A list of all protein sequences and the resulting clusters is available in Table S2 in the supplemental material.

Analysis of the Planctomycetes limnophilus core genome. Since P. limnophilus is the only planctomycete that can be genetically manipulated, we focused on its core genome in order to identify candidate proteins for subsequent genetic experiments. Consequently, all sequences except those with a P. limnophilus origin were deleted from the 114 core genome clusters obtained after the in silico subtraction of E. coli and B. subtilis genomes. The resulting clusters were manually inspected, and proteins likely to be involved in cell division, shape determination, or compartmentalization were selected as starting points for a “guilt-by-association” approach (see Table S3 in the supplemental material). The genomic context of each selected gene was manually inspected to identify putative operons (POs) harboring at least two members of the core genome. Five putative operons fulfilled these criteria and were manually analyzed by using BLAST, the NCBI database, and the Conserved Domain Database (CDD) for functional annotation (36).

Cell shape and cell division proteins. By screening the literature, we composed a list of 64 proteins involved in cell division, filament formation, and cell shape determination from various bacteria, archaea, and yeasts (see Table S4 in the supplemental material). The query protein was compared against the NCBI database by using BLAST with default settings and “Planctomycetes” as the taxon filter. Putative homologues were reverse analyzed against the database. If the reverse BLAST experiment revealed proteins from the same family of the original query, they are highlighted in green in Table S4 in the supplemental material; if not, they are highlighted in yellow and are referred to as “like” proteins. Protein identifications, E values, and percent identity values are given in Table S4 in the supplemental material.

Gene synteny and content of dcv operon genes among Planctomycetes. Based on previous studies, we used planctomycete proteins homologous to those of E. coli encoded within the dcv operon as a starting point (see Table S4 in the supplemental material (39, 46). Subsequently, we manually determined the synteny of the corresponding genes and visualized the gene order (see Fig. 4), while domains were determined as described above.

Analysis of planctomycetal signal transduction. The Microbial Signal Transduction (MiST) database (http://mistdb.com/) was employed to gain an overview of regulatory proteins encoded in the eight planctomycetal genomes (57). Extracytoplasmic function sigma factor (ECF) sequences were retrieved from the MiST database and cross-checked by an ECF finder analysis (http://ecf.jgi.doe.gov/), and all positive sequences were added to a comprehensive ECF data set that contained a total of 362 sequences (see Table S5 in the supplemental material). Subsequently, the protein sequences were aligned by using the ClustalW algorithm at the European Bioinformatics Institute website (http://www.ebi.ac.uk/). After the manual removal of gaps and N- or C-terminal extensions, phylogenetic distances were calculated for the resulting multiple-sequence-alignment files of the ECF core proteins by employing the Fitch-Margoliash and least-squares distance algorithms (9), as implemented in the BioEdit sequence alignment editor (18). Individual branches of the tree were subjected to detailed follow-up analyses, as described previously (53), to identify potential anti-σ factors and/or additional associated proteins. These analyses are based on genomic context conservation as a measure of a potential functional link between their respective gene products.

RESULTS AND DISCUSSION

The planctomycetal core genome. The Planctomycetes clade displays enormous phylogenetic depth, and its phylogeny is still con-
trouversial (23, 46, 59). Thus, prior to determining the planctomycete core genome, we used 16S rRNA gene analysis to decide which species to include in our study for the practical purpose of determining a group of organisms sufficiently related to each other to produce a meaningful core genome. The maximum likelihood 16S rRNA gene tree shows, with excellent bootstrap support (bootstrap value of 100), two distinct branches. One branch was comprised of anammox planctomycetes, while the other branch contained all other planctomycetal species cultivated thus far (Fig. 1). This result provided us with a phylogenetic criterion to leave anammox bacteria out of our analysis. Within this work, we thus differentiate between “Planctomycetes” and “anammox bacteria.” However, whether or not this has taxonomic implications is beyond the scope of this study.

Despite phylogenetic controversies, Planctomycetes are members of the domain Bacteria and, as such, would be expected to divide by using FtsZ. However, FtsZ is absent in planctomycetal genomes (2). In order to identify proteins that might be involved in cell division, shape determination, and compartmentalization, we performed two independent lines of analysis. Our planctomycetal core genome approach led to 114 predicted protein clusters containing 2,908 proteins from all eight analyzed planctomycetes after the in silico subtraction of E. coli and B. subtilis genomes (Fig. 2). Each cluster contained between 8 and 764 proteins, with an average of 25 proteins. This subtraction yielded important insights, since planctomycetes are difficult to cultivate, and genetic tools are available only for P. limnophilus. Since P. limnophilus is the only member of this phylum that has been genetically modified, we focused on its core genome for subsequent analyses. P. limnophilus contributed 274 proteins to the 114 core clusters, resulting in an average of about 2 proteins per cluster. However, even in this model, planctomycete mutant construction is rather slow. Thus, we wanted to further reduce the number of candidate genes for subsequent mutational studies. Consequently, all P. lim-

FIG 1 The class Planctomycetia is split into two distinct orders. Shown is a maximum likelihood phylogenetic tree of planctomycetal 16S rRNA gene sequences. Bootstrap values are shown only for deep branches, and different Escherichia coli sequences were used as an outgroup, represented by an arrow (GenBank accession numbers of the E. coli sequences used are GU594315, GU594305, GU594306, GU594304, GU594302, and GU594316). “Ca. Kuenenia stuttgartiensis,” “Candidatus Kuenenia stuttgartiensis.”

FIG 2 The planctomycetal core genome. The planctomycetal core genome was determined by comparing all eight sequenced Planctomycetes species against each other by using the BLAST algorithm with a coverage of >60% and an E value cutoff of 1e−5 as parameters. The Venn diagram visualizes the 564 clusters fulfilling these criteria. After in silico subtraction, using proteins encoded by Escherichia coli (GenBank accession number U00096) and Bacillus subtilis (accession number AL009126), 450 planctomycetal core genome clusters were eliminated, and 114 planctomycete-specific clusters remained. Genes within these 114 clusters fulfill two criteria: they are conserved among planctomycetes and absent in the genomes of the two most intensively studied model organisms, E. coli and B. subtilis.
nophilus core proteins were manually evaluated, and six clusters were selected for in-depth analyses based on their putative cell biological relevance (see Table S3 in the supplemental material). One of these clusters, cluster 23, contained 39 MC-like proteins encoded by all eight planctomycetal genomes under study, and 5 proteins belong to *P. limnophilus*. Despite the presence of MC-like proteins, no “smoking gun” was found, and given that most proteins are of unknown function, we followed a “guilt-by-association” approach to identify putative targets. First, we manually inspected the neighborhood within the six selected *P. limnophilus* core genome clusters. Second, we focused on genomic regions in which at least two core genes were situated within a putative operon. Five such POs fulfilled these criteria (PO1 to PO5) (Fig. 3), while MC-like proteins were not among them. However, the five putative operons encode proteins containing putative cell division-related domains (e.g., von Willebrand factor [vWF], PDZ, and ATPase domains). Such domains have been found to be associated with putative novel cell division mechanisms in *Archaea* (34, 35). Individually, each PO has features that render it attractive for further study.

In PO1, we identified forkhead-associated (FHA), SpoIIE, and LpxE domains. Such domains might be involved in cell division and membrane organization, as FHA domains play a critical role in the hyphal fusion of *Neurospora crassa*, while SpoIIE is required for asymmetric cell division during sporulation in *Bacillus* (37). LpxE domains are known to selectively dephosphorylate lipid A molecules, which are present in the outer membranes of Gram-negative bacteria (20). This finding is unexpected, since planctomycetes lack such an outer membrane. However, further genes required for the biosynthesis of lipid A were found in *Rhodopirellula baltica*, indicating that planctomycetes might produce lipid A (17, 24).

PO2 encodes membrane-bound dehydrogenase and heme-binding domains. This configuration is frequently found in *Planctomycetes* and *Verrucomicrobia* but is not found in any other bacteria sequenced thus far. Furthermore, dihydrolipoamide dehydrogenase and glycine cleavage H protein domains were identified. Most strikingly, some similarity to the MinD superfamily “P-loop ATPase ferredoxin” domains were detected. MinD participates in the spatial regulation of the formation of the cytoplasmic FtsZ ring and thus has a major function during cell division (41).

The third identified operon, PO3, encodes proteins which appear to be related to signal transduction. Two of them contain von Willebrand factor type A domains. Thus, they might be involved in processes such as membrane modeling, cellular differentiation, and adhesion. Genes encoding proteins of unknown function, such as Plim_0822, might be responsible for planctomycete-specific traits, and their being near the “usual suspects,” such as von Willebrand factor domains, makes them interesting targets for subsequent genetic experiments.

PO4 encodes 12 proteins. While two might be related to sugar metabolism, four are hypothetical or contain domains of unknown function. Thus, they might be involved in processes such as membrane modeling, cellular differentiation, and adhesion. Genes encoding proteins of unknown function, such as Plim_0822, might be responsible for planctomycete-specific traits, and their being near the “usual suspects,” such as von Willebrand factor domains, makes them interesting targets for subsequent genetic experiments.

PO5 is the largest putative operon, and its component proteins contain several ATPase domains. In addition, PDZ domains that may mediate protein-protein interactions are found together with trypsin domains, indicating putative proteases. Furthermore, a
cell adhesion domain found in bacteria (CARDB) was identified. In addition, the fact that three core genome members (Plim_1802 to Plim_1804) are next to each other makes PO5 a promising candidate for future investigations. For a more detailed description of the five putative operons, see the supplemental material.

As expected, the comparative genomic approach revealed several target proteins from P. limnophilus for future genetic analyses. Interestingly, while they are part of the core genome, MC-like-protein-encoding genes show a patchy distribution in the P. limnophilus genome and do not cluster with other core genome members. However, such proteins were recently demonstrated to be involved in the endocytosis of proteins in G. obscuriglobus and are thus, without a doubt, worth studying in detail (29, 48).

**Cell division and cell shape determination.** All eight planctomycete genomes used in this study were screened for genes encoding one of 64 cell morphogenesis- or division-related proteins that were described for different bacteria, archaea, or yeasts, and the results are summarized in Table S4 in the supplemental material. Table 2 provides a selection of proteins that were found in at least one planctomycete. As shape-determining proteins, we included representatives from all bacterial actin-like families, ParM, FtsA, Alp7, and MreB (49). The gene encoding the bacterial actin homologue MreB was found in Planctomyces species and in Blastopirellula marina. However, we failed to identify any bacterial tubulin (Btub)-like protein using BtubA and BtubB as well as FtsZ as queries (32, 45). Recently, a new family of FtsZ-like proteins (FtsZL-1) was described, and B. marina, G. obscuriglobus, Pirellula staleyi, and R. baltica were shown to encode such a protein (34). We failed to detect homologues of such FtsZL-1 proteins in all other sequenced planctomycetes using almost the same approach as that which led to the discovery of FtsZL-1 (34). In addition, FtsZL-1 proteins lack the T6-H7 structure and thus might not be able to form filaments at all (34). Thus, FtsZL-1 might very well fulfill a function other than Z-ring formation during cell division in planctomycetes. In addition, we screened for proteins involved in archaean cell division. Two FtsZ-independent division mechanisms are known for the domain Archaea (35). One is homologous to the eukaryotic ESCRT-III system and requires Snf and VPS4 proteins. While we could not detect the presence of Snf-encoding genes in planctomycetal genomes using BLAST (see Table S4 in the supplemental material), VPS4 revealed several putative homologues. However, a critical diagnostic feature, the amino-terminal microtubule interaction and transport (MIT) domain, present in all archaean VPS4 proteins, is missing in their putative planctomycetal counterparts. Thus, we conclude that planctomycetes lack VPS4. The third archaean division mechanism is still puzzling but seems to involve actin-like proteins. As mentioned above, by our search approach, MreB homologues could be found in only four planctomycetes and can thus not explain a universal mechanism of planctomycetal cell division. Finally, we did not detect homologues of the protein putatively involved in Z-ring formation in anammox bacteria, kustd_1438 (58). The absence of this protein among planctomycetes other than anammox bacteria further supports our initial assumption to focus on one part of the planctomycetal order, as two independent cell division mechanisms might further support the differences between both lineages. Since the cytokinesis of planctomycetes parallels the division of yeast, we screened for septin proteins known to be crucial for septum formation in Saccharomyces cerevisiae (3, 43). However, we did not identify any homologues among planctomycetes using the basic BLAST algorithm.

In contrast to putative Z-ring-forming proteins, FtsW was found in Planctomyces species, while genes for other cell division-related proteins, such as FtsE, FtsK, ParA, CpaE, EnvA, MraW, and MraY, were detected in all planctomycetal genomes. However, MraZ was found only in Planctomyces and Blastopirellula

---

**Table 2** Cell division- and cytoskeleton-related genes of planctomycetes

<table>
<thead>
<tr>
<th>Protein</th>
<th>Rhodopirellula baltica SH1</th>
<th>Planctomyces limnophilus</th>
<th>Pirellula staleyi</th>
<th>Isosphaera pallida</th>
<th>Planctomyces brasiliensis</th>
<th>Gemmata obscuriglobus</th>
<th>Planctomyces maris</th>
<th>Blastopirellula marina</th>
</tr>
</thead>
<tbody>
<tr>
<td>ClpX</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>ClpP</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>CpaE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>Ddl</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>EnA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FtsE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FtsL</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>FtsK</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>FtsW</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>RodA</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MraW</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MraY</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MraZ</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MreB</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MurC</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MurD</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MurE</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>MurF</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>MurG</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>ParA</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

a This table summarizes data shown in Table S4 in the supplemental material; only those proteins that are encoded by at least one planctomycetal genome are shown. Proteins are grouped by function, and boldface type indicates dcw operon-related proteins.
species. Besides cell division, “peptidoglycan-related” proteins such as MurC, MurD, MurE, MurF, and MurG as well as Ddl and FtsI showed a more complex distribution pattern. While all these proteins are encoded by \textit{Planctomyces} species, all except MurE and Ddl are absent in all other planctomycetes. Besides \textit{Planctomyces}, MurE was also found in \textit{B. marina}, \textit{R. baltica}, and \textit{P. staleyi}, while Ddl was absent in \textit{B. marina} (Table 2).

Beyond gene content, we investigated the gene synteny of cell morphogenesis and division genes. Here our starting point was the \textit{dcw} operon structure of \textit{E. coli} as a comparison. The \textit{dcw} operon is highly conserved among “rod-shaped” bacteria, and “genomic channeling” has been proposed to force the colocalization of genes involved in cell division and peptidoglycan synthesis to optimize septum formation (39). While genomic channeling nicely explained the persistence of the \textit{dcw} operon over a broad taxonomic range (39) (Fig. 4), this mechanism is supposed to be restricted to rod-shaped cells, and transitions to other morphologies seem to involve genomic rearrangements and a loss of gene order (54). This was supported by the first investigation of planctomycetal \textit{dcw} gene clusters (46). Despite the few planctomycetal genomes available at that time, the conclusion was drawn that the last planctomycetal ancestor (LPA) contained a complete \textit{dcw} operon and that gene content and order were gradually lost (46). This hypothesis is supported by our data shown in Table 2 and Fig. 4: all three \textit{Planctomyces} species contain most \textit{dcw} operon genes found in all planctomycetes. However, gene synteny is less conserved among \textit{Planctomyces} than gene content. While \textit{Planctomyces maris} has four putative operons and one single \textit{murD}-like gene, the members of the \textit{dcw} operon of \textit{E. coli} are scattered among eight different putative operons and one solitary \textit{murD} gene in \textit{Planctomycetes brasiliensis}. In \textit{P. limnophilus}, the situation is even more complex, as \textit{dcw} genes are localized across nine different putative operons. Among all other planctomycetes, \textit{P. staleyi} and \textit{R. baltica} contain four \textit{dcw} operon genes, while \textit{B. marina}, \textit{G. obscuriglobus}, and \textit{Isosphaera pallida} harbor three, none of which colocalize. Thus, gene synteny is less conserved than gene content in regard to planctomycetal \textit{dcw} genes, while the gradual loss becomes more obvious, since our study used more than double the number of planctomycetal genomes used in previous analyses (46). Interestingly, gene content and synteny reflect the 16S rRNA gene tree phylogeny (Fig. 1). However, the presence of peptidoglycan (PG)-related genes is difficult to explain, as planctomycetes are known to lack a PG cell wall (12), and genes without function tend to erode quickly. For PG-wall-less \textit{Chlamydia}, cell division has been shown to be altered by beta-lactam antibiotics. Consequently, speculations of an involvement of PG in Z-ring replacement arose (2). However, consistent with previous studies, we failed to detect any growth inhibition caused by beta-lactam antibiotics such as ampicillin among planctomycetes thus far (6, 22, 26).

Taken together, our finding that several cell division- and peptidoglycan-associated proteins known from bacteria are encoded by some, if not all, \textit{Planctomyces} (Table 2) might suggest that their last ancestor might have divided in an FtsZ-dependent manner and was confined by a PG layer.

**Putative cell division-related operons.** Following a guilt-by-association approach, we screened putative planctomycetal \textit{dcw} gene-containing operons for the presence of core genome members. The rationale driving this analysis is the assumption that cell division-related \textit{dcw} operon genes might be found near genes involved in the FtsZ-independent division of planctomycetes. Genes

---

**FIG 4** Content and synteny of \textit{dcw} operon genes among planctomycetes. The cell division-related (gray) and peptidoglycan synthesis-related (black) genes arranged in the \textit{dcw} operons of \textit{E. coli} and \textit{B. subtilis} are shown in comparison to homologous genes from planctomycetal genomes. Genes of unknown function are shown in white. Genes with weak similarity to the \textit{E. coli} query are labeled in gray (W, \textit{mraW}; Z, \textit{mraZ}; L, \textit{ftsL}).
encoding FtsZ-substituting proteins most likely arose in the LPA and would consequently be absent from genomes of other bacteria such as *E. coli* or *B. subtilis*. Such planctomycetal cell division genes are likely to be conserved among all planctomycetes and to be part of their core genome. However, only two putative operons fulfilled our screening criteria. One was found within the genome of *P. brasiliensis*, spanning 25 kb and containing 14 genes, 2 of which are *dcw* operon related (*ftsI* and *mraW*) and 4 of which are members of the planctomycetal core genome (*Plabr_0294, Plabr_0295, Plabr_0300, and Plabr_0301*) (Fig. 5). The second putative operon found in the genome of *I. pallida* had a size of more than 38 kb and contained 26 genes. Besides the *marY*-like *dcw* gene, three planctomycetal core genome members, *Isop_3026, Isop_3036*, and *Isop_3037*, were identified (Fig. 5). However, since both organisms are not genetically manipulable, we performed a second screening to compare the identified core genome members putatively related to cell division from *I. pallida* and *P. brasiliensis* operons against the genome of *P. limnophilus*. This screening revealed two and three putative *P. limnophilus* operons, respectively (Fig. 5; for a detailed description of individual genes and proteins, see the supplemental material).

Many proteins encoded by the genes of PO6 to PO10 contained domains putatively related to cell division, such as ATPase domains, or to chromosome segregation domains. In addition, many genes contain domains of unknown function that are found exclusively in Planctomycetes or Verrucomicrobia (Fig. 5). Most strikingly, three of four putative operons identified by the guilt-by-association approach turned out to contain two or three genes that are part of the planctomycetal core genome. In PO6 and PO7, two core genome members with identical domain architectures colocalize (*Plim_2793 and Plim_2794, and Plim_0775 and Plim_0776, respectively). Interestingly, the same domain architecture and localization pattern were observed in two areas of a *P. brasilensis* putative operon, where the first two genes, *Plabr_0294* and *Plabr_0295*, colocalize with *ftsI* and the other two, *Plabr_0300* and *Plabr_0301*, colocalize with *mraW*. This finding makes *P. limnophilus* PO6 and PO7 interesting candidates for future genetic studies.

All 10 identified putative operons (Fig. 3 and 5) provide some indications that their gene products are involved in cell shape determination, membrane organization, chromosome segregation, or cell division. The planctomycetal proteins of unknown function, containing known domains or not, might be the most interesting ones, as new biological functions might await elucidation. The rationale for our study was the *in silico* identification of putative targets for subsequent genetic experiments, and all the putative operons identified need future experimental verification to reveal their putative relevance in planctomycetal key traits. Wet laboratory experiments must now be carried out to determine whether colocalized genes are cotranscribed or not. Experimentally verified operons from the gene clusters identified in this study should then be subjected to subsequent mutagenesis in *P. limnophilus*. Such experimental approaches should help to ultimately identify the function of genes involved in planctomycete-specific traits.

**Planctomycetal signal transduction and gene regulation.** Complex life-styles and cellular differentiation cycles necessitate the presence of equally complex signaling cascades and regulatory networks to orchestrate and coordinate the corresponding massive spatiotemporal modulations of gene expression patterns, as

---

**FIG 5** Putative cell division-related planctomycetal operons. The putative Plabr and Isop operons fulfilled two criteria: (i) they contain a cell division- or peptidoglycan synthesis-related *dcw* gene, and (ii) at least one additional core genome member is present within the same putative operon. In a second round, corresponding *P. limnophilus* genes were identified, which turned out to be localized on 5 different putative operons (PO6 to PO10). Selected protein domains are highlighted, while core genome members are boxed in red.
has been observed for model organisms such as \textit{B. subtilis}, \textit{Caulobacter crescentus}, and \textit{Streptomyces coelicolor} (25, 30, 38). An understanding of the regulatory and signaling potential embedded in the genome sequences of \textit{Planctomycetes} species is therefore a prerequisite to the unraveling of differentiation processes in this phylum. We therefore decided to closely examine the repertoire of signal-transducing proteins of the planctomycete species.

Based on mechanistic principles and phylogenetic relationships, three fundamental mechanisms of bacterial signal transduction can be distinguished: one-component systems (1CSs), two-component systems (2CSs), and extracytoplasmic function sigma factors (ECFs) (52, 53, 56). Data based on raw-read-represented bacterial phyla such as the \textit{Proteobacteria} or the \textit{Firmicutes}, an “average” genome harbors about four times more 1CSs than 2CSs and four times more of the latter than ECFs. However, while the proportions of 1CSs and 2CSs are relatively constant for most phyla (56), the numbers of ECFs seem to be much more variable among different species (52). With very few exceptions, the absolute numbers of signaling proteins correlate directly with the size of the genome (56). The distribution of planctomycetal signaling proteins supports both the correlation between genome size and the number of signaling proteins as well as the ratio between 1CSs and 2CSs (Table 3). In contrast, all planctomycetes are particularly rich in ECFs, which are equal in number to or sometimes even outnumber the 2CSs, indicating an important role of these proteins in the regulatory processes of this phylum.

1CSs. 1CSs are proteins that connect an input with an output on a single polypeptide chain. The majority (about 85%) of 1CSs from most bacteria harbor a DNA-binding output domain, as is the case for paradigms such as LacI or TetR. The output of the remaining 15% is mostly realized using protein kinases/phosphatases or regulators involved in the making and breaking of secondary messengers, such as adenylate/diguanlylate cyclases (56). In planctomycetes, the percentage of DNA-binding 1CSs is much lower, while 20 to 45% of all 1CSs are Ser/Thr protein kinases (STPKs) (Table 3). Such proteins, which are common among eu-karyotes, normally play a minor role in bacterial signal transduction, and their physiological role has only very recently been elucidated for a few examples, most of which are involved in differentiation or developmental processes (44). The reason for the very different abundances of STPKs between prokaryotes and eu-karyotes might reside in the fact that in bacteria, with their chromosome located in the cytoplasm, most adaptation and differentiation processes are directly regulated via differential transcription. For this, bacteria use predominantly sensory histidine kinases (HKs) as part of 2CSs to activate their cognate response regulators (RRs), which, once activated, function mostly as DNA-binding proteins. In contrast, the eu-karyotic chromosomes are spatially separated from the cytoplasm in the nucleus. Hence, direct protein modifications (including phosphorylations) play a much more important role in cellular physiology and differentiation. Thus, eu-karyotes require larger numbers of protein kinases to modulate the protein phosphorylation pattern in response to diverse stimuli. Considering that the chromosome of, e.g., \textit{Gemmatia obscuriglobus} is surrounded by two nuclear membranes, the high abundance of STPKs might point toward a similar role of protein phosphorylation in the differentiation processes of planctomycetes (12). However, this does not imply any necessary evolutionary homology but implies merely a convergence of signal transduction system characteristics due to analogies in cell organization.

2CSs. 2CSs consist of a sensor protein, the HK, which, in the presence of a suitable stimulus, autophosphorylates. Subsequently, a cognate partner protein, the RR, becomes active through a phosphoryl group transfer from the histidine kinase (HisKA) domain of the HK to the receiver domain of the RR (15). This phosphorylation then activates the output domain of the RR, which, in most cases (about 85%), binds DNA to orchestrate the differential expression of its target genes (14, 56). Usually, the two partner proteins are encoded by neighboring genes that are organized in an operon. In planctomycetal genomes, the absolute number of 2CS proteins correlates well with genome sizes, and the ratio of 1CSs to 2CSs is comparable to those in other species (Table 3). However, RRs outnumber HKs, and only 50 to 70% of the RRs are DNA-binding proteins (bacterial average, 85%). Most planctomycet al RRs consist of a receiver domain only and lack obvious output domains. Many planctomycetal 2CS sensor proteins represent complex (hybrid) HKs, containing multiple putative input domains and often more than one HisKA domain. Hybrid HKs are part of complex phosphorelays, involving the integration/modulation of the HK activity itself or the activity of downstream RRs (7). The majority (about 70 to 80%) of HKs and
RRs are encoded by orphan genes unrelated to their respective partners, pointing toward a more complex mode of 2CS-dependent signal transduction in planctomycetes. Among such genes, one was found to encode a novel type of RR conserved in all eight planctomycetal genomes but restricted to this phylum. Its domain architecture consists of an N-terminal Ser/Thr/Tyr protein kinase domain ( Pfam designation, Pkinase) and a C-terminal receiver domain (REC) domain (Fig. 6A), which is normally found in the N terminus. The widely distributed domain architecture of RRs consisting solely of a receiver domain might point toward a role of these proteins in the shuttling of phosphoryl groups between histidine kinases and histidine-containing phosphotransfer proteins (containing a Pfam HPT domain) within complex 2CS-based phosphorelays involved in integrating different stimuli, as described previously for the decision-making process in B. subtilis sporulation or the Rcs phosphorelay of enterobacteria (8, 33, 60). This hypothesis seems attractive in light of the greater abundance of RRs than HKs. A second explanation could be that such “receiver-only” RRs mediate their output through direct protein-protein interactions, as most classically exemplified by the CheY-like RRs involved in the chemotaxis systems discussed below. Taken together, only a small fraction of planctomycetal 2CSs provides a simple one-to-one connection between a stimulus input and a cellular response in the form of differential gene expression. Instead, the majority of 2CS proteins seem to be involved in complex, highly interconnected, multistep phosphorelays required to orchestrate some aspects of the complex life cycle described for this phylum. Interestingly, the overall situation is very reminiscent of that of the deltaproteobacterium Myxococcus xanthus (50). However, the unusual domain configuration of some planctomycetal RRs (Fig. 6A) remains enigmatic. Probably, upon the phosphorylation of the receiver domain by a cognate HK, the N-terminal protein kinase domain is activated to phosphorylate its target protein(s) as an output of 2CS-mediated signal transduction. While both the physiological role and the signaling mechanism remain elusive, the conservation of these novel RRs within all planctomycetal species and their absence outside this phylum make these proteins very interesting candidates for future functional studies.

Chemotaxis. While bacterial chemotaxis is derived from 2CSs, we treat such proteins as a separate entity, because of the large number of unique domains and proteins dedicated to orchestrating bacterial motility. All planctomycetes included in this study are motile in one stage of their life cycle, while I. pallida lacks a flagellum, and its motility is achieved through gliding (16). Thus, chemotaxis is to be expected for all the analyzed bacteria. We focused on the distribution and abundance of two central chemotaxis proteins, so-called methyl-accepting chemoreceptor proteins (MCPs) and the CheA-like HK, in order to gain some insight into the motility intelligence between different species. While the first gives an indication of the diversity of different stimuli sensed, the second is a reliable measure of the number of independent chemotaxis systems encoded in a given bacterial genome. Two major conclusions can be drawn from the data shown in Table 3. First, there is a huge variance in the complexity of the chemotaxis systems among the eight species, which is mostly independent of genome size. While B. marina, P. staley, and especially G. obscuriglobus have a high motility intelligence, with four to five independent chemotaxis systems and 20 to 60 MCPs, the remaining planctomycetes contain only a very limited number of chemotaxis proteins, based on one or two chemotaxis modules. Second, our analysis revealed the complete lack of any chemotaxis machinery in the motile and flagellated organism R. baltica, which is in agreement with data from previous studies (17). This can mean that R. baltica can swim but has no means to direct its movement, or it somehow has acquired, developed, or adapted a completely different mechanism for adjusting its flagellar motor. Both alterna-
Fig. 7 Phylogenetic tree and classification of ECF sigma factors from planctomycetes. Shown is a phylogenetic tree of all ECF sequences extracted from the MiST2 database, generated by the least-squares distances method. A subsequent group analysis of unclassified planctomycetal ECFs was done based on their phylogenetic distances, genomic context conservation, or the presence of promoter motifs, which resulted in the assignment of the novel groups ECF45, ECF46, ECFSTK1, ECFSTK2, ECFSTK3, and ECFSTK4 as well as the ECF01-like groups ECF01-Gob and ECF01-P. The corresponding sequence information is shown in Table S5 in the supplemental material. Different groups are shown in different colors and are highlighted correspondingly in the tree map. (A) Abundance and distribution of ECFs in the different planctomycete species. Abbreviations: Rba, Rhodopirellula baltica SH1; Pat, Pirellula staleyi DSM 6068; Pma, Planctomyces maris DSM 8797; Pli, Planctomyces limnophilus DSM 3776; Pbr, Planctomyces brasiliensis DSM 5305; Ipa, Isovista pallida ATCC 43644; Gob, Gemmata obscuriglobus UQM 2246; Bna, Blastopirellula marina DSM 3645. (B) Phylogenetic tree of ECFs in planctomycete species. Eight ECFs were excluded because of their phylogenetic distance from other ECFs. These proteins are highlighted in Table S5 in the supplemental material.

...tives seem equally unlikely but point toward very interesting future experiments in this planctomycetal model organism.

ECFs. ECFs belong to the pool of alternative σ factors that recognize specific promoter sequences to control gene expression. Their activity is usually regulated by cognate anti-σ factors (5, 19). Recently, a classification of ECFs was reported, based on about 400 microbial genomes (53). Given the genome sequencing bias toward model bacteria, or those with medical or biotechnological relevance, many phyla were underrepresented despite their ecological relevance. Not surprisingly, most ECFs from the dominant phyla can be classified, while a majority of ECFs from underrepresented phyla, including the Planctomycetes, could not be assigned to any defined ECF group (53). Because of the large abundance of unclassified ECFs in Planctomycetes, we performed an in-depth analysis of these proteins mimicking the initial ECF classification (53). The overall abundance of ECFs in the eight sequenced planctomycetal species is shown in Fig. 7A. Branches of the phylogenetic tree corresponding to known or novel conserved ECF groups are color coded in Fig. 7B, and their genomic abundance and distribution are indicated by the same colors in Fig. 7A. ECFs were found to represent a widely distributed signaling principle for the phylum Planctomycetes. G. obscuriglobus in particular is one of the most ECF-rich bacteria sequenced to date, with its genome encoding 115 ECFs. This is only three proteins short of the current record holder, the deltaproteobacterium Plesiocystis pacifica. Based on the original set of 43 group-specific hidden Markov models (HMMs) described previously by Staron et al. (53), only 96 out of the 362 proteins could be classified into one of the four conserved ECF groups present in planctomycetes (data not shown). The majority of these belong to the highly diverse ECF01 group of RpoE-like σ factors that are functionally linked to typical membrane-anchored anti-σ factors harboring an antisigma domain. The other classifiable ECFs were assigned either to the two poorly characterized groups ECF22 and ECF42, which both lack a designated anti-σ factor, or to the more distantly related ECF-like proteins of the ECF43 group, which seem to be functionally linked to STPKs (53). In this new planctomycete-specific analysis, we defined eight novel ECF groups based on phylogenetic relationships and genomic context conservation, raising the count of classified planctomycetal ECFs by a factor of 3, from 96 to 304 (Fig. 7; see also Table S5 in the supplemental material for details). Most strikingly, the ECF01-Gob branch (Fig. 7B, light green) contains more than half of all ECFs encoded in the genome of G. obscuriglobus (62 of 115) but not a single protein from any of the other seven planctomycetal species. This strongly suggests that more than 50% of all ECFs in this organism are evolutionarily young paralogues. In addition, these Gemmata-specific ECFs comprise a diverse and unusual domain architecture (Fig. 6B). They share very large C-terminal extensions, sometimes exceeding the length of the ECF core by a factor of more than 4 (19). The extension of about 20 ECF01-Gob proteins contains multiple WD40 repeats that might be involved in the assembly of multiprotein complexes (51). In addition, most ECF01-Gob proteins are membrane anchored, harboring 1 to 3 putative transmembrane helices between the ECF core and the C-terminal extension. These are the first membrane-anchored ECFs described for bacteria. Given the presence of a nucleus-like cellular organization in this organism, it is a highly attractive, but also purely speculative, hypothesis that this unusual type of transmembrane ECF has specifically evolved in this species to facilitate signal transduction between the cytoplasm and the nucleoid.

Besides ECF01-Gob, we identified four well-separated novel ECF groups (ECFSTK1 to ECFSTK4) that are all functionally linked to STPKs, according to genomic context conservation.
Such a link was previously described only for the ECF-like proteins of the ECF43 group (53). This result demonstrates the wide distribution of signaling modules consisting of a protein kinase and an ECF-type σ factor. While it is tempting to speculate that, in these modules, the STPKs function as sensors that phosphorylate and thereby activate their cognate ECF partner, such a novel mechanism of ECF-dependent signal transduction awaits experimental verification.

Taken together, our ECF analysis of the phylum Planctomycetes indicates that the exploration of this third pillar of bacterial signal transduction has only just begun. Especially in phyla underrepresented in the data set of the initial analysis (53), there is a large potential for finding novel conserved mechanisms not present in the more common model organisms. The groups described above not only add to the growing list of conserved ECFs but also allow a first glimpse into how planctomycetes employ such proteins as a central mechanism of signal transduction.

**Conclusion.** By employing various bioinformatic methods, we identified proteins putatively involved in planctomycetal cell division and in their conspicuous cell plan. In addition, we identified planctomycetal signal transduction systems as well as sigma factors and evidence for a new signaling logic among Planctomyces. As the genetic manipulation of Planctomycetes remains tedious, our findings will help to guide future wet laboratory experiments with a manageable list of the most promising target proteins to study hallmark planctomycetal traits.

**REFERENCES**


