HHS Public Access

Author manuscript

Nat Microbiol. Author manuscript; available in PMC 2021 January 01.

Published in final edited form as:

Nat Microbiol. 2020 January ; 5(1): 126–140. doi:10.1038/s41564-019-0588-1.

Cultivation and functional characterization of 79 planctomycetes uncovers their unique biology

Sandra Wiegand[1], **Mareike Jogler**[2], **Christian Boedeker**[2], **Daniela Pinto**[3], **John Vollmers**[4], **Elena Rivas-Marín**[5], **Timo Kohn**[1], **Stijn H. Peeters**[1], **Anja Heuer**[2], **Patrick Rast**[2], **Sonja Oberbeckmann**[6], **Boyke Bunk**[2], **Olga Jeske**[2], **Anke Meyerdierks**[7], **Julia E. Storesund**[8], **Nicolai Kallscheuer**[1], **Sebastian Lücker**[1], **Olga M. Lage**[9], **Thomas Pohl**[10], **Broder J. Merkel**[10], **Peter Hornburger**[2], **Ralph-Walter Müller**[11], **Franz Brümmer**[11], **Matthias Labrenz**[6], **Alfred M. Spormann**[12], **Huub J. M. Op den Camp**[1], **Jörg Overmann**[2], **Rudolf Amann**[7], **Mike S. M. Jetten**[1], **Thorsten Mascher**[3], **Marnix H. Medema**[13], **Damien P. Devos**[5], **Anne-Kristin Kaster**[4], **Lise Øvreås**[8], **Manfred Rohde**[14], **Michael Y. Galperin**[15], **Christian Jogler**[1,16,*]

[1]Radboud University Nijmegen, The Netherlands. [2]Leibniz Institute DSMZ, Braunschweig, Germany. [3]TU Dresden, Dresden, Germany. [4]Karlsruhe Institute of Technology, Karlsruhe, Germany. [5]Centro Andaluz de Biología del Desarrollo (CABD)–CSIC, Pablo de Olavide University, Seville, Spain. [6]Leibniz Institute for Baltic Sea Research Warnemünde (IOW), Rostock, Germany. [7]Max Planck Institute for Marine Microbiology, Bremen, Germany. [8]University of Bergen, Bergen, Norway. [9]University of Porto, Porto, Portugal. [10]TU Bergakademie Freiberg, Freiberg, Germany. [11]University of Stuttgart, Stuttgart, Germany. [12]Stanford University, Stanford, CA, USA. [13]Wageningen UR, Wageningen, The Netherlands. [14]HZI Braunschweig, Braunschweig, Germany. [15]NIH, Bethesda, MD, USA. [16]Friedrich Schiller University Jena, Jena, Germany

Abstract

When it comes to the discovery and analysis of yet uncharted bacterial traits, pure cultures are essential as only these allow detailed morphological and physiological characterization as well as genetic manipulation. However, microbiologists are struggling to isolate and maintain the majority of bacterial strains, as mimicking their native environmental niches adequately can be a challenging task. Here, we report the diversity-driven cultivation, characterization and genome sequencing of 79 bacterial strains from all major taxonomic clades of the conspicuous bacterial phylum Planctomycetes. The samples were derived from different aquatic environments but close relatives could be isolated from geographically distinct regions and structurally diverse habitats, implying that ‘everything is everywhere’. With the discovery of lateral budding in ‘Kolteria novifilia’ and the capability of the members of the Saltatorellus clade to divide by binary fission as

*Correspondence to: christian@jogler.de.

Author contributions

S.W., M.J. and C.J. designed the study. M.J., T.K., A.H., P.R., J.E.S., O.M.L. and L.Ø. cultivated the planctomycetes and established axenic cultures. M.J., C.B., T.K., A.H., P.R., S.O., O.J., J.E.S., T.P., B.J.M., P.H., R.-W.M., F.B., M.L., A.M.S., A.-K.K., L.Ø., A.M. and C.J. were involved in the sampling, sample processing and basic enrichments. S.W., J.V. and B.B. sequenced and assembled the genomes. E.R.-M. and D.P.D. constructed the deletion mutants, A.M. and R.A. constructed the planctomycetal fosmid, S.L. obtained the planctomycetal MAGs and O.M.L. isolated DNA for sequencing. C.B., T.K., S.H.P., M.R. and C.J. performed the microscopy. H.O.d.C., M.S.M.J., T.M., J.O., M.H.M. and R.A. provided expertise and supervision. S.W., D.P., J.V., N.K., M.H.M., M.Y.G. and C.J. analyzed and interpreted data. S.W. and C.J. wrote the manuscript with major contributions from M.J., C.B., D.P., J.V., S.O., N.K., R.A., M.H.M, L.Ø. and M.Y.G., and with help and approval from all authors.

well as budding, we identified previously unknown modes of bacterial cell division. Alongside unobserved aspects of cell signalling and small-molecule production, our findings demonstrate that exploration beyond the well-established model organisms has the potential to increase our knowledge of bacterial diversity. We illustrate how 'microbial dark matter' can be accessed by cultivation techniques, expanding the organismic background for small-molecule research and drug-target detection.

One Sentence Summary

Planctomycetes unveil novel features of regulation, small molecule production, morphology, cell division and amoeba-like locomotion.

Introduction

Most of the knowledge on bacterial key features, such as cell biology, molecular genetics, metabolism, virulence and biotechnological exploitability, has been gained from studying model organisms of the phyla Proteobacteria, Firmicutes, Actinobacteria and Bacteroidetes. However, as about 80 bacterial phyla have been recognized, does the exploration of model bacteria indeed provide a comprehensive representation of general bacterial traits and concepts? With the advent of next-generation sequencing and metagenomics, it became possible to investigate the genomes of uncultivated bacteria from sparsely studied phyla. Many differ tremendously from the current models and comparative analyses of metagenome-assembled genomes (MAGs) provide insights into their unexplored potential. However, detailed studies of unseen physiological or cell biological properties and unknown biochemical features necessitate axenic cultures. From a curiosity-driven perspective, this raises the question of what we can learn about the nature surrounding us by a more detailed exploration of under-sampled bacterial phyla?

To address this issue, we focused on one of the most enigmatic but low-represented phyla, the Planctomycetes[1], which comprise only 29 described genera, most of them represented by only one species. Planctomycetes have been shown to be phylogenetically distinct and to cluster together with the biotechnologically and medically important phyla Verrucomicrobia and Chlamydiae within the PVC superphylum[2]. The ubiquitous planctomycetes play an important role in the global carbon and nitrogen cycles[1], are of biotechnological relevance for wastewater treatment[3] and are suspected to produce bioactive small molecules[4]. Although some of the most astonishing planctomycetal traits, such as their putative ancestry to the eukaryotes, have been recently questioned[5–8], planctomycetes undoubtedly feature unique cell biology attributes[1], such as multiplication without the otherwise universal bacterial cell-division protein FtsZ[1]. Members of the class Phycisphaerae and the family *Candidatus* 'Brocadiaceae' both divide by binary fission and 'Brocadiaceae' strains possess an intracytoplasmic membrane forming the anammoxosome[3]. On the other hand, species belonging to the class Planctomycetia divide by budding, which occurs either polarly in Planctomycetaceae or arbitrarily in the round cells of the Gemmataceae and Isosphaeraceae. Their cytoplasmic membrane can be intensively invaginated, forming an enlarged periplasmic space putatively used for the digestion of macromolecules[7]. Such macromolecules are assumed to be internalized by crateriform structures and their associated

fibres[7]. These doughnut-shaped pits in the outer membrane of the Planctomycetia are a unique hallmark trait of this class. Furthermore, Planctomycetia form stalks and perform a lifestyle switch from sessile mother to swimming daughter cell[1,9–11].

We employed a straightforward cultivation procedure along with diversity-driven strain selection to gain further insights into the planctomycetal biology. This ‘deep-cultivation’ approach focused on selectively obtaining axenic cultures from all known clades of the Planctomycetes. Our endeavour was complemented by genome sequencing and analysis, combined with thorough microscopic characterization. The application of this combined approach to the under-sampled bacterial phylum Planctomycetes resulted in an unprecedented overview presented in this ‘umbrella’ publication. We unearthed unexpected bacterial traits and demonstrated that far more is ‘out there’ in terms of cell biology and small-molecule biosynthesis than the canonical model bacteria suggest.

Results and Discussion

Sampling, cultivation, phylogeny and central genomic characteristics of 64 previously unknown planctomycetal species -

Planctomycetes are known to be particularly abundant in aquatic ecosystems (for a review see Wiegand and colleagues[1]). Hence, we based our sampling strategy on this knowledge and followed the hypothesis that highly diverse habitats are likely to harbor a wide range of divergent Planctomycetes. Therefore, we sampled various natural as well as man-made aquatic environments (Fig. 1).

A carefully developed enrichment strategy, favoring the generally slow growth of Planctomycetes and their preference for *N*-acetylglucosamine[12], was employed to target the Planctomycetes inhabiting either the sampled water column, sediments, iron-hydroxide deposits or the biofilms formed on the sampled surfaces. Despite the method-inherent selection for aerobic, heterotrophic, neutrophilic and mesophilic bacteria, we were able to obtain a total of 79 strains in axenic cultures (Fig. 1 and Supplementary Table 1) in addition to four MAGs from enrichment cultures and an 81 kb fosmid assembly from the samples from the Black Sea.

To test our initial hypothesis, we assessed the phylogenetic diversity of the obtained isolates and elucidated their occurrence and distribution among metagenomic samples (Fig. 1a). With this strategy we were able to isolate planctomycetes from all of the main taxonomic clades and from many branches that have previously only been known from the gene analyses of 16S rRNA. Most isolates cluster within the Planctomycetaceae, a clade known to harbor predominately marine isolates[1]. However, as most of our samples are of marine origin, this result is not surprising. Interestingly, we found several members of the same proposed species – for example, V7, Mal65 and Pan14r – to originate from diverse habitats, such as microbial mats, beach sediments and a seawater aquarium. It became obvious that this ubiquity seems to be a common trait of the planctomycetes when we searched for the corresponding 16S rRNA sequences with >99% sequence identity in the Sequence Read Archive database. The majority of samples derived from seawater or living in association with a host organism have been found (with an abundance >0.005%) in both water and host-

derived samples. Together, these findings point to Beijerinck's principle of 'everything is everywhere'[13] and might indicate that diversity in local ecological niches, rather than geography, might allow to further increase the diversity of axenic planctomycetal strains in the future. On a side note, at least some such promising niches might be located in the Baltic Sea. Not only did 20 of our isolates, including our deepest branching one, originate from this brackish sea, but it was also the source of a vast, but unpublished, strain collection established by Heinz Schlesner in the 1980s.

We determined the genome sequences of all 79 isolated strains along with the genome of one previously described strain. Furthermore, we resequenced seven formerly analyzed strains (Fig. 2, green font) as their available genome sequences were of low quality. Altogether, the genomes of 62 strains could be closed, whereas the remaining 27 genomes have a median scaffold count of 23.

Multiple tree variations were calculated, including genomes publicly available before May 2017 to unravel the taxonomic position of the 89 sequenced genomes within the planctomycetal phylum. The high quality of the gained genomic information allowed multilocus sequence analyses (MLSA; Supplementary Fig. 1 and Supplementary Data 2–9), with additional trees (Supplementary Data 10–19) and composite supertrees to be computed based on gene content, amino acid identity, ribosomal proteins and RpoB sequences as well as 16S rRNA marker genes (Supplementary Fig. 2). Despite the MLSA-based phylogeny contradicting the 16S rRNA gene-based tree in several branches, this approach in combination with the additional trees allowed the expansion of the known 16S rRNA gene-based planctomycetal taxonomy lineages[14] to a tree based on whole-genome information. This tree holds the *Pirellula*, *Bythopirellula* (which also comprises the Pir4 lineage[15]) and *Gimesia* clades as well as the Gemmataceae and Isosphaeraceae, all of which are members of the class Planctomycetia. It also includes the Phycisphaerae and the 'Brocadiaceae'. In addition, by culturing members of the deep-branching *Saltatorellus* clade – a taxon without any axenic cultures thus far – we were able to unearth previously unexplored branches of the planctomycetal phylogenetic tree (Fig. 2). Based on the 16S rRNA[16] (Supplementary Table 2), average nucleotide identity, amino acid identity and morphological parameters, we assigned the 79 isolated strains to 1 known and 65 previously unknown species belonging to 8 known and 31 previously unknown genera (Fig. 2). Although in-depth taxonomy and valid species descriptions are beyond the scope of this study, we propose tentative names for all isolates that will guide future valid descriptions of these and related isolates (Fig. 2 and Supplementary Table 1).

The sequenced genomes were between 3.3 and 11.0 Mb (median 7.2 Mb) in size. However, this broad range narrows down for the specific clades. Although 'Brocadiaceae' and Phycisphaerae have smaller genomes, the Gemmataceae, Isosphaeraceae and the *Pirellula* clade trend towards big genomes (Fig. 2 and Supplementary Fig. 3). Between one and five plasmids were identified in six of the closed genomes. Based on high-quality genomes, the core genome of the planctomycetal phylum comprises only 92 genes (Supplementary Figs. 4,5 and Supplementary Tables 3, 4), echoing the broad diversity in planctomycetal cell biology and lifestyle. We conducted a COG analysis to obtain an initial insight into the general metabolism of the isolates (Supplementary Fig. 6). The Gemmataceae and

Isosphaeraceae seem to have large proportions of genes dedicated to transcription and signal transduction, whereas the Isosphaeraceae and the *Saltatorellus* clade share a high number of genes involved in cell-wall/membrane/envelope biogenesis. A general trend towards the presence of genes related to the transport and metabolism of carbohydrates, amino acids and inorganic ions was observed. However, the question of what potential may be hidden in the huge genomes of the planctomycetal phylum remains to be properly addressed, given that between 40 and 50% of the predicted proteins are annotated as hypothetical, a number that is considerably higher than for most other bacterial lineages[17,18].

In the following three sections, we summarize our key findings based on the obtained 79 isolates and 89 sequenced genomes.

Planctomycetes challenge the concepts of bacterial cell division and shape determination.

In contrast to studies based solely on omics, we were able to employ (time-lapse) light and electron microscopy to investigate the cell division and morphology of our axenic cultures (Fig. 3). As expected, Planctomycetia divide by budding (Fig. 3b(i)–(vi)). Among the axial symmetric pear-shaped cells, such as *Planctopirus limnophila*, daughter cells always originate from the same bulge region opposite to the pole that forms the holdfast structure or stalk19 (Fig. 3b(i)–(vi)). Radial symmetric/coccoid cells also perform budding. Strikingly, our isolate '*Kolteria novifilia*' Pan216, which forms a distinct phylogenetic clade (Fig. 3a), deviated from that pattern: its daughter cells originated at a mid-cell position, perpendicular to the longitudinal axis of the mother cell (Fig. 3b(vii),(viii) and Supplementary Movie 1). We suggest the term 'lateral budding' for this hitherto unknown bacterial cell division mechanism.

In contrast to budding Planctomycetia, coccoid 'Brocadiaceae' and Phycisphaerae divide by binary fission. No obvious morphological differences to canonical coccoid Gram-negative bacteria were observed (Fig. 3b(v),(vi)). Surprisingly, the mostly coccoid cells from the deep-branching *Saltatorellus* clade can shift between budding and binary fission (Fig. 3b(ix)–(xii) and Supplementary Movie 2), a behavior that has not been seen for any bacterial clade thus far.

Although the budding process as such is not understood at the molecular level, binary fission has been well studied. In free-living bacteria, a conserved set of 12 essential proteins forms the divisome[20], the bacterial tubulin-homolog FtsZ being the key component[1,21]. Our in-depth genomic analysis verified the previously proposed absence of ftsZ[22,23] in all analyzed planctomycetal genomes, including the deep-branching *Saltatorellus* clade (Fig. 3a, Supplementary Fig. 7 and Supplementary Table 5). Some strains encode an FtsZ-like protein (Fig. 2, Circle 13). However, such FtsZl-1 proteins are unlikely to form filaments22. Although all planctomycetes encode the divisome proteins FtsE and FtsK, FtsI and FtsW were only found in some species (Fig. 3a, Supplementary Fig. 7 and Supplementary Table 5). Thus, FtsI and FtsW might not be essential for planctomycetal division, a hypothesis that we verified by constructing and analyzing *P. limnophila ftsI* and *ftsW* deletion mutants (see Methods). As *ftsX* is absent in all planctomycetal genomes, the encoded ATPase FtsE most probably does not perform peptidoglycan remodelling during cell division because it requires interaction with FtsX[24] (Fig. 3a, Supplementary Fig. 7 and Supplementary Table 4).

The DNA translocase FtsK is also encoded by all planctomycetes. Given that we could not construct an *ftsK*-deletion mutant, it seems to be essential and could serve as a starting point to reveal the molecular background of the planctomycetal cell division in the future.

As previously assumed[22], there was no correlation between the presence/absence of canonical cell division genes and the mode of planctomycetal cell division, that is, polar, lateral and arbitrary budding or binary fission (Fig. 3a, Supplementary Figs. 8,9 and Supplementary Table 4). This further indicates that the canonical bacterial cell-division proteins (except FtsK) might not relate to planctomycetal reproduction.

In addition to cell division, planctomycetal cell shape determination is also an enigmatic process. In the absence of a static rigid cytoskeleton, bacterial shape is usually defined by the peptidoglycan sacculus[21]. We recently demonstrated experimentally that, contrary to the previous expectations[25], the main planctomycetal clades possess a canonical peptidoglycan cell wall[5,6]. However, by employing a more detailed and stringent computational analysis than previously used[5,6], we found that not all planctomycetes use the canonical set of genes to build peptidoglycan (Fig. 2, Supplementary Fig. 10 and Supplementary Table 4). For example, *Gemmataceae* and members of the *Pirellula* clade seem to lack most peptidoglycan-synthesis genes, despite possessing a generic peptidoglycan cell wall[5,6].

Although peptidoglycan determines the cell shape, its making and breaking is controlled by cytoskeletal filaments such as the actin homolog MreB. MreB is found in the majority of sequenced bacterial species, where it organizes the elongasome multi-enzyme complex that directs peptidoglycan synthesis in canonical elongated species such as *Escherichia coli* and *Bacillus subtilis*[21,26–28]. It is assumed that it does so in the majority of bacteria26. However, the rod-like-shaped strain '*Kolteria novifilia*' and the other *Planctomycetia* lack MreB, whereas the coccoid *Phycisphaerae* and '*Brocadiaceae*' as well as members of the *Saltatorellus* clade encode MreB (Fig. 3a). Some rod-shaped bacteria, such as members of *Rhizobiales*, *Actinomycetes* and *Firmicutes* also lack MreB[26], and functions other than cell-shape determination have been described for MreB in *Helicobacter pylori*[29] and chlamydiae[30–32]. Because chlamydiae are closely related to planctomycetes but require MreB for their cell division[30], it was tempting to test whether MreB is also involved in planctomycetal cell division. As anecdotal observations pointed towards inhibition of cell division of *P. limnophila* by the MreB inhibitor A22, we constructed a *P. limnophila ΔmreB* mutant that divided like the wild type in time-lapse microfluidic experiments and showed no phenotypic difference compared with the wild type (Methods). Thus, the MreB in planctomycetes is probably not involved in either cell division or in cell-shape determination. Given the sometimes very unusual morphology and polymorphisms that we observed among our isolates (Supplementary Figs. 11–19), the planctomycetal mechanisms of shape determination might differ from those in model bacteria.

In conclusion, most canonical essential bacterial proteins related to cell-shape determination and division as well as peptidoglycan synthesis are either absent in planctomycetes or seem to fulfil different functions. Whether planctomycetes employ non-canonical enzymes for peptidoglycan biosynthesis and cell division or their enzymes are evolutionarily so distinct that they escaped our computational detection remains elusive. However, our determination

of the non-essentiality of otherwise universally essential genes, such as *ftsI*, points towards the former. Planctomycetes hence challenge the universality of the concepts deduced from typical model organisms and necessitate a broader view on bacterial cell biology. This endeavour will benefit from the axenic cultures obtained in this work.

Signaling capacity of planctomycetes

To determine the signal transduction preferences of the phylum *Planctomycetes*, we investigated the occurrence of one- and two-component systems, extracytoplasmic function (ECF) σ factors, chemotaxis signalling and the secondary messenger c-di-GMP[33–35] (Fig. 4). As the number of encoded signal-transduction systems unsurprisingly increases with genome size (Supplementary Fig. 20), it became evident that no generalized phylogenetic pattern exists and instead different planctomycetal clades encode different types, and especially different numbers, of signal transduction components (Fig. 4 and Supplementary Table 6).

The larger the genome, the greater the number of histidine kinases and response regulators that together form the two-component systems (Supplementary Figs. 21,22). Accordingly, the members of the two clades with the largest genome sizes, *Gemmataceae* and *Isosphaeraceae*, demonstrated the highest count, whereas the '*Brocadiaceae*' and *Phycisphaerae* had significantly fewer two-component systems. This pattern also included hybrid histidine kinases (Supplementary Fig. 23), which enable more complex phosphorelays by phosphorylating their own internal receiver domain. If the number of response regulators is plotted against the number of histidine kinases, one would expect a ratio of approximately 1:1, which is usually observed in bacteria[36]. However, the planctomycetes had distinctly more response regulators than histidine kinases (Supplementary Fig. 24), suggesting that they have many divergent two-component systems where one histidine kinase is able to phosphorylate multiple response regulators.

Similarly, the number of predicted Ser/Thr/Tyr protein kinases and phosphoprotein phosphatases generally correlated with the genome size (Supplementary Fig. 25), at least for members of the *Planctomycetia* and the *Saltatorellus* clade. In contrast, '*Brocadiaceae*' and *Phycisphaerae* possess only a few such enzymes, independent of their genome size. This is similar to our observations for chemoreceptors (methyl-accepting chemoreceptor proteins, MCPs), as their occurrence also does not relate to genome size (Supplementary Fig. 26).

In addition, the number of diguanylate cyclases and phosphodiesterases that are involved in the synthesis and degradation of c-di-GMP is largely independent of the genome size (Supplementary Fig. 27). The largest numbers were found in the '*Brocadiaceae*' and the *Pirellula* clade, whereas *Gemmataceae* and *Isosphaeraceae* generally had fewer such enzymes despite their huge genomes. Whether this finding is related to the biofilm-associated lifestyle would require a more detailed analysis.

Furthermore, we addressed the question of the quantity and nature of the ECF σ factors in planctomycetes (Fig. 4). Whereas the '*Brocadiaceae*' strains almost entirely lack ECFs, members of the *Saltatorellus* clade, the *Isosphaeraceae* and the *Gemmataceae* are extraordinarily rich in ECF-encoding genes (Supplementary Fig. 28 and Supplementary

Table 7). These taxa carry ten or more ECFs per megabase, in contrast to other bacteria, which on average have four ECFs per megabase[37]. A total of 62 ECF groups have been described before[22,38,39], ten of which are exclusively restricted to planctomycetes[22,40] (Supplementary Table 8). In this study, we describe 30 previously unknown groups (Supplementary Fig. 29 and Supplementary Table 9). Extracytoplasmic function groups usually represent known regulatory mechanisms, for example, mediated by membrane-bound or soluble anti-sigma factors, C-terminal extensions or phosphorylation[41,42]. Surprisingly, some of the previously uncharacterized groups (ECF87–89) comprise ECFs that have elongated N-termini, which have not been observed before (Supplementary Fig. 30). They have weak homology with known transcriptional regulators and are rarely found outside of the planctomycetes (Supplementary Figs. 31,32). These N-termini might represent additional DNA-binding domains possibly generated by the fusion of a transcriptional regulator with the ECF, a hypothesis worthy of future investigation.

The capability of Planctomycetes to produce (novel) small molecules

Planctomycetes have already been considered to be untapped ‘talented’ producers of small molecules with a potentially therapeutic character4,12,43,44. As these analyses were based on a far smaller dataset, we chose to reiterate the capacity of planctomycetes to produce small molecules. We found all analyzed planctomycetal genomes to encode between 2 and 15 biosynthetic gene clusters (BGCs; Fig. 5, Supplementary Fig. 33 and Supplementary Table 10). The identified BGCs are predicted to be involved in the production of polyketides and non-ribosomal peptides as well as terpenes, bacteriocins, ectoines, ladderanes and others (Supplementary Figs. 34,35). The former are of special interest, as many known natural bioactive secondary metabolites – for example, antibiotics – are either produced by polyketide synthases (PKS) or by non-ribosomal peptide synthases (Supplementary Figs. 36–44). Members of the *Pirellula* clade and *Isosphaeraceae* (Supplementary Figs. 36–40,43) harbor three to six gene clusters encoding putative PKSs, whereas typically no more than four PKS clusters were identified in members of the other taxa. One exception was *Schlesneria paludicola* in the *Gimesia* clade, which harbors as many as seven putative PKS-encoding gene clusters (Supplementary Fig. 42). The overall abundance of putative gene clusters encoding non-ribosomal peptide synthases is lower compared with PKSs, is highly restricted to the *Pirellula* and *Bythopirellula* clades and reaches up to four in ‘*Rosistilla oblonga*’ CA51 (Supplementary Figs. 33–41). The *Gimesia* clade, in turn, comprises most of the clusters putatively involved in the production of bacteriocins and ectoines, both of which seem to be barely present in the *Pirellula* and *Bythopirellula* clades (Supplementary Fig. 37). This gives the impression that not all planctomycetal clades are equally prone to produce small molecules. In particular, members of the *Saltatorellus* and ‘*Brocadiaceae*’ clades, and the *Phycisphaerae* contain fewer BGCs than other planctomycetal strains (Fig. 5).

We performed a sequence-similarity network analysis to investigate whether the identified planctomycetal BGCs differ from those of other bacteria. In a previous study covering the bacterial domain, it was shown that 72% of all BGCs found are connected to one component, whereas only 7.6% do not have any connection with other BGCs[45]. Planctomycetes did not emerge in this latest global analysis due to the lack of available genomes[45]. Our network analysis joined planctomycetal BGCs (Supplementary Fig. 45,

colored dots) with BGCs of known function obtained from the MIBiG database[46] (black dots). Among the 265 identified clusters and 628 singleton hits, 59 and 56% are derived from planctomycetal BGCs. With the default cutoffs of BiG-SCAPE[47], only 1.1% of the clusters contain MIBiG as well as planctomycetal BGCs (Fig. 5). These results show that the vast majority of putative planctomycetal BGCs are not connected to known BGCs and thereby indicate an untapped metabolic potential, especially of members of the *Planctomycetia*. The missing links of BGCs of the same group and different planctomycetal clades (Fig. 5 and Supplementary Figs. 36–44) provide support for the idea that planctomycetes might be specialized producers achieving specific activities of the produced compounds, depending on their lifestyle in terms of environmental conditions and the presence of competing bacteria.

The most talented producers known thus far are the actinobacteria, which sometimes have more than 40 BGCs[45]. However, the genome-mining algorithms48 that were employed for that analysis were trained on well-described BGCs, which in turn were derived from small-molecule-producing bacteria, such as myxobacteria and actinobacteria. Therefore, direct comparisons of this number with the significantly lower numbers found in planctomycetes might be misleading, as a proportion of planctomycetal BGCs could have been missed. One future source of unknown small molecules could lie in a deeper analysis of the giant genes featured by many planctomycetes[49] (Supplementary Fig. 46). As we have shown, planctomycetes lack a canonical cell division machinery. This should make them intrinsically resistant to antibiotics targeting that machinery, which they indeed are. Together with the capacity of their large genomes, this makes them ideal candidates for the production of yet unknown antibiotics by hitherto unknown biosynthetic pathways.

Summary and Outlook

In this study we isolated and sequenced 79 planctomycetal strains. We were able to gain insights into the diversity of this phylum with regards to signalling, potential of small molecule production and cell-division strategies. In light of our findings, it is questionable whether the available axenic cultures of canonical (model) organisms reflect bacterial diversity sufficiently, as the planctomycetes are dramatically different from model bacteria in fields that will need more attention in future studies.

We have shown that the culturing of highly diverse bacteria is indeed possible when their ecological niche is sufficiently well mimicked and time is allowed for slow-growing strains to propagate. The application of similar cultivation strategies to other understudied bacterial phyla will increase the number and phylogenetic diversity of axenic cultures in the future. There are most probably many more microorganisms ‘out there’ worth exploring. Although we can gain valuable insights from metagenomics or single-cell techniques, an axenic culture is ultimately needed for complete understanding of the physiology of an organism. However, simple discovery-driven culturing might not serve this purpose. Thus, we assert the need for deep cultivation: a hypothesis and diversity-driven holistic approach, combining the appropriate cultivation techniques with genomic and morphological analyses well beyond the classical ways of describing previously uncharacterized species.

Materials and Methods

Sampling

Details on the sampling of each strain are summarized in Supplementary Table 1. The isolation of planctomycetes was performed using multiple techniques, including solid-agar or gellan-gum plates, liquid enrichment cultures and floating filters, as described in this section. For cultivation on agar or gellan-gum plates (see below for the composition of the medium), the sampled material originating from Panarea, Monterey Bay, Heligoland, Mallorca, Corsica and a seawater fish tank was washed with sterile seawater containing 20 or 100 mg l^{-1} cycloheximide or 20 ml l^{-1} nystatin suspension (Sigma-Aldrich) to prevent fungal growth. Afterwards, the sampled material was swabbed over plates and placed on the middle of plates containing different variations of carbon/nitrogen sources, antibiotics and antifungal agents (see below). In addition, biofilm suspensions were prepared by carefully scraping off the algal/seaweed biofilms into sterile natural seawater using single-use scalpels for samples from Heligoland and Corsica. Plates were inoculated with 20 µl of the biofilm suspension at different dilutions (0–10^{-2}). Sediment samples from Panarea, the seawater fish tank and Mallorca were inoculated with 100 µl homogenized sample material at dilutions of between 0 and 10^{-2} per plate. The isolates that originated from pond-water samples were obtained by plating 20 or 100 µl of homogenized water samples. The freshwater sponge samples were washed with sterile filtered water from their surroundings and sponge pieces were swabbed over plates. The microplastic and wood particles collected on 4 September 2014 were incubated for 16 days at the Baltic Sea coast as previously described by Oberbeckmann and colleagues[50]. The samples were stored for eight weeks at 4°C before bacterial biofilm isolation by β-galactosidase digestion (2 mg ml^{-1}; 28 units ml^{-1}) for 30 min at 30°C and simultaneous vortexing every 5 min followed by sonication for 10 min at 30°C. The subsequent separation of plastic or wood particles from the biofilm was performed by filtration. The biofilms were stored for 11 months at 4°C before plating. The incubated plastic particles collected on 8 October 2015 (incubation time of two weeks) were washed three times with sterile natural seawater and stored at 4°C until cultivation (5 days after sampling). The plastic particles were vortexed and 50 µl seawater was used for cultivation on agar or gellan-gum plates. In addition, the plastic particles were placed directly on solid plates. Liquid enrichment cultures were used to isolate planctomycetes from various habitats (Heligoland, Panarea and Monterey Bay). For the enrichment cultures from Heligoland, 100 ml medium (see below for the composition of the medium) was inoculated with a piece of algae (3 cm in diameter) prewashed with sterile seawater containing cycloheximide (20 or 100 mg l^{-1}). In addition, 100 ml medium was inoculated with 100 µl of the biofilm suspension. For the enrichment cultures originating from Monterey Bay, different pieces of kelp (leaf, buoyancy elements, stem or holdfast structure) were placed in 100 ml medium. The material from the Panarea samples was directly inoculated in 100 ml liquid medium. After several weeks of incubation, the enrichment cultures were checked for enrichment of planctomycetes by wide-field microscopy. Characteristic traits, such as polar budding or pear-to-spherical cell shapes, provided an indication of the possible presence of planctomycetes and the cultures were plated on solid medium. The floating-filter technique (adapted from Sipkema *et al.*[51]) was used for Heligoland samples by filling six-well plates with 5 ml medium containing antibiotics and

fungicide. For inoculation, the algae biofilm suspensions were diluted 1:5, 1:10 or 1:50 in 5 ml sterile seawater and filtered onto black polycarbonate filters (0.1 μm retention size, 25 mm in diameter). The filters were placed floating on the medium using sterile tweezers and incubated at 20°C until colony growth was observed. The colonies were then transferred to solid plates. All strains were identified by Sanger sequencing of the 16S rRNA genes; on confirmation of their planctomycetal identity, single colonies were re-plated at least three times for purification and resequenced (for a detailed description see Rast and colleagues[52]).

Media and cultivation conditions used for isolation of planctomycetes

In general, the composition of the media used was inspired by previous work by Schlesner[53] and Lage and Bondoso[54] on the cultivation of planctomycetes. All media for marine as well as limnic planctomycetes contained 2.38 g l^{-1} HEPES as a buffer and 20 ml l^{-1} mineral salt solution with 10 g l^{-1} nitrilotriacetic acid, 29.7 g l^{-1} $MgSO_4 \cdot 7H_2O$, 3.34 g l^{-1} $CaCl_2 \cdot 2H_2O$, 0.01267 g l^{-1} $Na_2MoO_4 \cdot 2H_2O$, 0.099 g l^{-1} $FeSO_4 \cdot 7H_2O$ and 50 ml l^{-1} metal salt sol. 44. The nitrilotriacetic acid was dissolved in 700 ml distilled water by adjusting the pH to 7.2 with KOH. All further components were dissolved separately and added slowly. The solution was sterilized by filtration and stored at 4°C. Metal salt sol. 44 consists of 250 mg l^{-1} Na_2–EDTA, 1,095 mg l^{-1} $ZnSO_4 \cdot 7H_2O$, 500 mg l^{-1} $FeSO_4 \cdot 7H2O$, 154 mg l^{-1} $MnSO_4 \cdot H_2O$, 39.5 mg l^{-1} $CuSO_4 \cdot 5H_2O$, 20.3 mg l^{-1} $CoCl_2 \cdot 6H_2O$ and 17.7 mg l^{-1} $Na_2B_4O_7 \cdot 10H_2O$. In the first step, EDTA was dissolved and, if required, a few drops of concentrated H_2SO_4 were added to retard precipitation of the heavy metal ions. The solution was sterilized by filtration and stored at 4°C. Furthermore, each medium was supplemented with 5 ml l^{-1} vitamin solution consisting of 10 mg l^{-1} *p*-aminobenzoic acid, 4 mg l^{-1} biotin, 20 mg l^{-1} pyridoxine hydrochloride, 10 mg l^{-1} thiamine hydrochloride, 10 mg l^{-1} Ca-pantothenate, 4 mg l^{-1} folic acid, 10 mg l^{-1} riboflavin, 10 mg l^{-1} nicotinamide and 0.2 mg l^{-1} vitamin B12. The *p*-aminobenzoic acid was dissolved first; the solution was sterilized by filtration and stored in the dark at 4°C. In addition, 1 ml l^{-1} trace element solution (1.5 g l^{-1} Na–nitrilotriacetate, 500 mg l^{-1} $MnSO_4 \cdot H_2O$, 100 mg l^{-1} $FeSO_4 \cdot 7H_2O$, 100 mg l^{-1} $Co(NO_3)_2 \cdot 6H_2O$, 100 mg l^{-1} $ZnCl_2$, 50 mg l^{-1} $NiCl_2 \cdot 6H_2O$, 50 mg l^{-1} H_2SeO_3, 10 mg l^{-1} $CuSO_4 \cdot 5H_2O$, 10 mg l^{-1} $AlK(SO_4)_2 \cdot 12H_2O$, 10 mg l^{-1} H_3BO_3, 10 mg l^{-1} $NaMoO_4 \cdot 2H_2O$ and 10 mg l^{-1} $Na_2WO_4 \cdot 2H_2O$) was added to each medium. The solution was sterilized by filtration and stored in the dark at 4°C. HEPES and mineral salt solution were added before autoclaving, and the vitamin and trace element solutions were added after autoclaving. The marine medium was further supplemented with 250 ml l^{-1} concentrated artificial seawater (46.94 g l^{-1} NaCl, 7.84 g l^{-1} Na_2SO_4, 21.28 g l^{-1} $MgCl_2 \cdot 6H_2O$, 2.86 g l^{-1} $CaCl_2 \cdot 2H_2O$, 0.384 g l^{-1} $NaHCO_3$, 1.384 g l^{-1} KCl, 0.192 g l^{-1} KBr, 0.052 g l^{-1} H_3BO_3, 0.08 g l^{-1} $SrCl_2 \cdot 6H_2O$ and 0.006 g l^{-1} NaF) before autoclaving. Strains originating from the Heligoland samples were initially isolated on solid and liquid media containing 80% natural seawater instead of artificial seawater. For carbon and nitrogen sources, the initial cultivation medium for marine planctomycetes was supplemented with 0.25 g l^{-1} glucose, 0.25 g l^{-1} Bacto peptone (BD Biosciences), 0.25 g l^{-1} BD Bacto yeast extract (BD Biosciences) and with or without 1 g l^{-1} N-acetylglucosamine.

Medium containing 1 g l^{-1} N-acetylglucosamine as the sole carbon source was also used. The medium containing N-acetylglucosamine as a sole carbon source in combination with

antibiotic agents and antifungal agents (see below) was most efficient for the isolation of marine bacteria. A medium containing 0.25 g l^{-1} glucose, 0.25 g l^{-1} Bacto peptone (BD Biosciences), 0.25 g l^{-1} BD Bacto yeast extract (BD Biosciences) and 1 g l^{-1} N-acetylglucosamine was used for the maintenance of marine strains. The only exception was strain MalM25, which grew only with 1 g l^{-1} N-acetylglucosamine as the sole carbon source. Peptone and yeast extract were autoclaved with the medium, glucose and N-acetylglucosamine were added afterwards from filter-sterilized stock solutions. For the initial isolation of limnic planctomycetes, the medium was supplemented with 0.25 g l^{-1} glucose, 0.25 g l^{-1} Bacto peptone (BD Biosciences) and 0.25 g l^{-1} BD Bacto yeast extract (BD Biosciences) with or without 1 g l^{-1} N-acetylglucosamine as well as 1 g l^{-1} N-acetylglucosamine as the sole carbon source. In addition, medium supplemented with 250 ml l^{-1} concentrated artificial seawater, 0.25 g l^{-1} glucose, 0.25 g Bacto peptone (BD Biosciences), 0.25 g l^{-1} BD Bacto yeast extract (BD Biosciences), 1 g l^{-1} N-acetylglucosamine and antibiotics (see below) was used for the isolation of the strains ETA_A1 and ETA_A8. For the maintenance of limnic planctomycetes, the medium was supplemented with artificial freshwater (final concentration in the medium: 0.53 mg l^{-1} NH_4Cl, 1.4 mg l^{-1} KH_2PO_4, 10 mg l^{-1} KNO_3, 49.3 mg l^{-1} $MgSO_4$·7H2O, 14.7 mg l^{-1} $CaCl_2$·$2H_2O$, 25 mg l^{-1} $CaCO_3$ and 25 mg l^{-1} $NaHCO_3$) before autoclaving. Furthermore, 0.25 g l^{-1} glucose, 0.25 g Bacto peptone (BD Biosciences), 0.25 g l^{-1} BD Bacto yeast extract (BD Biosciences) and 1 g l^{-1} N-acetylglucosamine were added to the medium. The peptone and yeast extract were autoclaved with the medium and the glucose and N-acetylglucosamine were added afterwards from filter-sterilized stock solutions. The pH of the media for the initial cultivation of marine and limnic planctomycetes was adjusted to a pH approximating that of the habitat (ranking between 6.5–8.0). For maintenance, the media were adjusted to pH 7.5 for the marine strains and to pH 7.0 for the limnic strains. Furthermore, the initial cultivation experiments were incubated at 20 °C or 28 °C, depending on the habitat water temperature. To prevent fungal growth, 20 mg l^{-1} cycloheximide and/or 20 ml l^{-1} nystatin (Sigma-Aldrich) were added to the media. In addition, different antibiotic combinations and concentrations favouring planctomycetal growth were tested. As an optimal standard mixture in marine media, 100 mg l^{-1} carbenicillin, 200 mg l^{-1} ampicillin and 500 mg l^{-1} streptomycin were identified to allow good selection of planctomycetes. Streptomycin concentrations above 100 mg l^{-1} inhibited bacterial growth in the limnic media. Ampicillin was used at a concentration of 200 mg l^{-1} and carbenicillin at 100 mg l^{-1} and 2,000 mg l^{-1}. As 2,000 mg l^{-1} carbenicillin completely inhibited bacterial growth if added before the plate pouring, a 100 µl stock solution (100 mg ml^{-1}) was applied directly on the plates before inoculation. All antibiotic concentrations still allowed the growth of various bacteria and no clear selection for planctomycetes was detected. For the isolation of strains ETA_A1 and ETA_A8, the medium contained 200 mg l^{-1} ampicillin and 100 mg l^{-1} carbenicillin. Most isolates were obtained on solid media containing either 15 g l^{-1} washed (three times with deionized water) agar (Bacto, BD) or 8 g l^{-1} gellan gum (Gelrite, Serva). Both solidifying agents were autoclaved separately and added to the medium before pouring the plates. Strains V22, V7 and Pan14r only grew on gellan gum (but not on agar plates), whereas strain Pan265 only grew in liquid medium.

Isolation and cultivation of the strains TBK1r, K2D and SV_7m_r

The strains K2D and TBK1r were isolated from iron-hydroxide deposits obtained at a depth of 1,734 m from the Valu Fa Ridge in the Pacific Ocean14. In brief, strain K2D was obtained from an inoculum in M30 medium53 with 70% aged seawater and 200 mg l^{-1} ampicillin for enrichment culturing. The samples were added at ratios of 1:10 and 1:100, and incubated at room temperature. The enrichments were examined through light microscopy after one week. Following this, 1:100 enrichments were plated onto M30 medium gelrite plates with 70% aged seawater, 200 mg l^{-1} streptomycin and 200 mg l^{-1} ampicillin. Small colorless colonies were observed after two weeks and the strain was purified by continuous re-streaking on M30 gelrite plates. The TBK1r strain was obtained from an iron-hydroxide dilution series in PBS (tenfold dilution series up to $1:10^4$). A 100-µl volume of these dilutions was then directly plated onto seawater–peptone–yeast extract–gelrite plates with 200 mg l^{-1} streptomycin and 200 mg l^{-1} ampicillin. Pink colonies appeared after two weeks and these were examined using light microscopy and streaked on new plates; the colonies displaying budding growth or rosette formation were re-streaked. Strain SV_7m_r was isolated from a meromictic lake in Bergen, Norway. Liquid M30 medium was inoculated with meromictic lake water from a depth of 7 m (at a dilution of 1:10 and 1:100) and incubated at room temperature. After an incubation period of three weeks, 10 µl of the enrichment was plated onto solid M13 gelrite plates with 200 mg l^{-1} ampicillin. Small red colonies were observed after three weeks and the strain was purified by continuous re-streaking on M13 gelrite plates.

Light microscopy

Phase-contrast analyses were performed employing a Nikon Eclipse Ti inverted microscope with a Nikon DS-Ri2 camera. Specimens were immobilized in MatTek glass-bottom dishes (35 mm, no. 1.5) employing a 1% agarose cushion[7]. Images were analyzed using the Nikon NIS-Elements software (Version 4.3). To determine the cell size, a minimum of 100 representative cells were counted either manually or using the object-count tool (smooth: 4×, clean: 4×, fill holes: on, separate: 4×). For time-lapse microscopy, cells were imaged either on a media-supplemented cushion[55] or on the CellASIC ONIX microfluidic platform (Merck Millipore) employing a Nikon Eclipse Ti inverted microscope with a Nikon DS-Ri2 camera. To image the strain Poly30 on a media-supplemented cushion, 1% agar or 1% agarose pads supplemented with the corresponding growth medium were used to immobilize the cells in a MatTek glass-bottom Microwell dish (35 mm dish, 14 mm Microwell with a no. 1.5 cover glass P35G-1.5–14-C). The cells were monitored for up to 30 h at the corresponding incubation temperature. A constant flow of medium was provided for at least 5 min at 5 psi to image strain Pan216 in the CellASIC ONIX microfluidic platform (Merck Millipore). The cells were then introduced and elastically trapped to the flow chamber using the manufacturer’s loading protocol (CellASIC). The cells from strain Pan216 were monitored at a constant flow of the corresponding medium for up to 21.5 h at the corresponding growth temperature. Images were aligned and analyzed using the NIS-Elements software V4.3 (Nikon Instruments).

Electron microscopy

For field emission scanning electron microscopy, bacteria were fixed in 1% formaldehyde in HEPES buffer (3 mM HEPES, 0.3 mM $CaCl_2$, 0.3 mM $MgCl_2$ and 2.7 mM sucrose, pH 6.9) for 1 h on ice and washed once with the HEPES buffer. Cover slips with a diameter of 12 mm were coated with a poly-L-lysine solution (Sigma-Aldrich) for 10 min, washed in distilled water and air dried. The solution of fixed bacteria (50 µl) was placed on a cover slip and allowed to settle for 10 min. The cover slips were then fixed in 1% glutaraldehyde in TE buffer (20 mM Tris and 1 mM EDTA, pH 6.9) for 5 min at room temperature and subsequently washed twice with TE buffer before dehydrating in a graded series of acetone (10, 30, 50, 70, 90 and 100%) on ice for 10 min at each concentration. Samples from the 100% acetone step were brought to room temperature before placing them in fresh 100% acetone. The samples were then subjected to critical-point drying with liquid CO_2 (CPD 300, Leica). The dried samples were covered with a gold/palladium (80/20) film by sputter coating (SCD 500, Bal-Tec) before examination in a field emission scanning electron microscope (Zeiss Merlin) using the Everhart Thornley HESE2 detector and the in-lens SE detector in a 25:75 ratio at an acceleration voltage of 5 kV.

Sequencing

The DNA from the planctomycetal strains and from samples of an anaerobic lab-scale bioreactor was extracted and purified using the Genomic-tip 100/G kit (Qiagen). The protocol was performed as recommended by the manufacturer with one exception – the incubation time with the digestive enzymes was prolonged to an overnight step. An additional bead-beating step was performed for some strains to ensure complete lysis of the cells. DNA from the hypoxic lab-scale sequencing batch reactor was extracted as described by van Kessel and colleagues[56]. Paired-end sequencing libraries were prepared using the TruSeq DNA PCR-free LT library preparation kit (Illumina) according to the manufacturer's instructions for the Low Sample (LS) protocol for 550 bp libraries in the reference guide. The prepared samples were quantified using the NEBNext library quant kit for Illumina (New England Biolabs). Mate-pair sequencing libraries were made with the Nextera mate pair library preparation kit (Illumina) following the manufacturer's gel-plus protocol using gel size selection in the reference guide. The libraries were sequenced on an Illumina MiSeq system employing 150 or 600 cycles with a MiSeq reagent kit v3 (Illumina). For strains K22.7, FF011L, K23.9 and HG15A2, libraries for short-read whole-genome sequencing were prepared with the TruSeq DNA sample prep kit v2 (Illumina) and sequencing was performed on the HiSeq 2500 (Illumina) for 100 cycles in both directions. SMRTbell template libraries were prepared according to the instructions from Pacific Biosciences, following the procedure and checklist for the preparation and sequencing of greater than 10 kb templates. Briefly, 8 µg of the genomic DNA libraries were sheared using g-tubes for the preparation of 10 kb libraries (Covaris). The DNA was end-repaired and ligated overnight to hairpin adaptors applying components from the DNA/Polymerase binding kit 2.0 (Pacific Biosciences). BluePippin (Sage Science) size selection to 4 kb (7 kb for strains K22.7, FF011L, K23.9 and HG15A2) was performed according to the instructions from the supplier. The conditions for the annealing of the sequencing primers and binding of polymerase to the purified SMRTbell template were assessed with the Calculator in RS Remote (Pacific Biosciences). Single-molecule real-time (SMRT) long-read sequencing was

carried out on the PacBio RSII taking 240-min videos for each SMRT cell utilizing P6 chemistry. The video length was 180 min for strains K22.7, FF011L, K23.9 and HG15A2. In the case of strain FF011L, nine SMRT cells were run applying the P4 chemistry. For strains K22.7, HG15A2 and K23.9, P5 chemistry in combination with one or two SMRT cells per strain was used.

Pre-assembly processing

For both types of Illumina sequencing reads, adaptor clipping and read trimming was done with Trimmomatic v0.35 (ref.[57]). For the mate-pair data, virtual read libraries of paired-end and mate-pair reads were created with NxTrim v160227 (ref.[58]). After adaptor removal, FastQ Screen v0.4.4 (ref. 59) was employed to filter the datasets for contamination and reads of low complexity were discarded using PRINSEQ lite v0.20.4 (ref.[60]). Subsequently, overlapping paired-end reads were merged using FLASH v1.2.11 (ref.[61]).

Assembly

Genomes with sequencing data from solely paired-end libraries or with data from paired-end and mate-pair libraries were assembled using SPAdes v3.7.0 (ref.[62]), Velvet v1.2.10 (ref.[63]) and an approach in which the output of the Velvet assembly was used as ‘untrusted-contigs’ input for a SPAdes assembly (see Supplementary Table 1).

Long-read genome assembly was performed using the RS_HGAP_Assembly.3 protocol SMRT Portal version 2.3.0, applying standard parameters (version 2.2.0 was used for the strain FF011L). The final contigs were error-corrected with all of the obtained Illumina reads by employing BWA[64] and subsequent variant and consensus calling with VarScan v2.3.7 (ref.[65]) and GATK[66].

In some cases (see Supplementary Table 1), where HGAP assembly performance was not sufficient, the PacBio and Illumina paired-end sequencing reads were assembled with the hybridSPAdes algorithm of SPAdes v3.7.0 (refs.[62,67]).

Scaffolding

Scaffolding was performed in several steps, not all of which were applied to all genomes, as specified in Supplementary Table 1. The first step was generally the scaffolding of pre-assembled contigs based on the reuse of paired-end and mate-pair libraries using SSPACE[68], followed by an approach to close the formed gaps with Sealer[69]. A subsequent approach was to achieve better scaffolds by adding the information of genome sequences of closely related organisms through the usage of the MeDuSa scaffolder[70]. For some genomes, a PCR-based approach led to the determination of gap sizes, and a subsequent Sanger-sequencing approach[71] was used to resolve the full sequence of some of these gaps.

Post processing

The gained consensus sequences were trimmed, circularized and adjusted to dnaA as the first gene. All genome assemblies were then examined using CheckM[72] and QUAST[73]. Some of the genomes (see Supplementary Table 1) were manually cured by employing the Integrative Genomics Viewer[74] following Bowtie 2 mapping[75]. Finally, genome annotation

was performed with Prokka v1.11 (ref.[76]). All genomes were submitted to the NCBI GenBank and are available at the accession numbers provided in Supplementary Table 1.

Fosmid sequencing

Fosmid library construction from microbial mats of the anoxic part of the Black Sea[77] has been described by Meyerdierks and colleagues[78]. Fosmid insert end sequences were determined for most of the clones[79]. In addition, the libraries were screened for the presence of bacterial 16S rRNA genes using the primer set GM3 and GM4 (ref.[80]). One of the clones carrying the 16S rRNA gene of an uncultured planctomycetal bacterium (000404d10) and a clone (000412e07) that overlapped it were completely sequenced as described by Meyerdierks and colleagues[78]. The two fosmids exhibited an overlap of 4,301 bp with no nucleotide polymorphisms and thus could be assembled to a contiguous sequence of 81.2 kb (fos2004AM). The contig (Supplementary Table 1) was annotated as described above with Prokka v1.11 (ref.[76]).

Sequencing and assembly – Phycisphaerae bins RAS1 and RAS2

The two phycisphaeral genomes RAS1 and RAS2 were gained by differential coverage binning of the metagenomic sequencing data of a bioreactor that was inoculated with biomass from a recirculating aquaculture system biofilter. A complete description of the cultivation, sequencing and subsequent analyses has been published by van Kessel and colleagues[56].

Analysis of the 16S rRNA of the complete phylum Planctomycetes

The planctomycetal 16S rRNA alignment was obtained from the non-redundant SILVA SSU Ref NR 99 database, release 13282 (ref.[81]). In addition, 16S rRNA sequences from the genomes reported in this study were aligned with SINA[82] and appended. The phylogenetic analysis was performed using FastTree v2.1.10 (ref.[83]) with the GTRGAMMA substitution model invoked. The tree was collapsed and formatted using iTOL v4 (ref.[84]).

MLSA

Multilocus sequence analysis was performed with all of the genomes reported in this study and all planctomycetal genomes available from the NCBI and IMG/M85 in April 2017 with a contamination rate <10% (ref.[72]). The completeness[72] of the genomes was either >90% or >95% (c90 or c95). To achieve maximum comparability and reduce bias caused by divergent open reading frame-calling methods, all publicly available reference genomes were re-annotated together with the genomes reported in this study, using Prokka v1.11 (ref.[76]). Orthologous groups for all genomes were detected based on a best bidirectional BLAST approach using Proteinortho[86]. The e-value threshold was set to 1×10^{-8}, the identity threshold to 30%, the query coverage was either >50% or >70% (cov50 or cov70) and the 'selfblast' option was enabled. The resulting pan-genome matrix was used to determine the single copy core genome. The resulting homologs were aligned with MUSCLE v.3.8.31 (ref.[87]), reduced with Gblocks v0.91b[88] and then concatenated. The different phylogenetic analyses (Supplementary Table 11) were done using either a maximum-likelihood (ML) approach (250 or 1,000 bootstraps, model of amino acid substitution:

PROTGAMMAIWAG) or Bayesian interference (BI; rate variation model: invgamma, amino acid model: fixed (wag), single chain per analysis, number of generations: 100,000 to 1,500,000)[89,90]. Two verrucomicrobial genomes served as the outgroup. The resulting trees are given in Newick format in Supplementary Data 2–9.

16S rRNA of isolates

The 16S rRNA gene phylogeny was computed with all of the 16S rRNA gene sequences from genomes reported in this study, all planctomycetal genomes available from the NCBI and IMG/M[85] in April 2017 with a full available 16S rRNA gene sequence and all SILVA entries originating from a validly described species (April 2017; 182 genomes in total). The 16S rRNA gene sequences were aligned with SINA[82]. The phylogenetic analysis was done using either an ML approach (1,000 bootstraps, model of nucleotide substitution: GTRGAMMAI) or BI (MrBayes default parameters, number of generations: 2,500,000)[89,90]. Five verrucomicrobial 16S rRNA genes served as the outgroup. The resulting trees are in Newick format in Supplementary Data 10,11 and Supplementary Fig. 2. A similarity matrix of the 16S rRNA gene identities is provided in Supplementary Table 2.

RpoB

Phylogeny for the RpoB protein sequence was computed with all of the genomes reported in this study and all planctomycetal genomes available from the NCBI and IMG/M[85] in April 2017 with a full available rpoB gene (182 genomes in total). The RpoB sequences were aligned using MUSCLE v.3.8.31 (ref.[87]) and reduced with Gblocks v0.91b[88]. The phylogenetic analysis was done using either an ML approach (1,000 bootstraps, model of amino acid substitution: PROTGAMMAIWAG) or BI (rate variation model: invgamma, amino acid model: fixed(wag), single chain per analysis, number of generations: 2,500,000)[89,90]. Two verrucomicrobial RpoB sequences served as the outgroup. The resulting trees are in Newick format in Supplementary Data 12 and 13.

Ribosomal proteins

The ribosomal protein phylogeny was computed for the 15 ribosomal proteins (L2, L3, L4, L5, L6, L14, L15, L16, L22, L24, S3, S8, S10, S17 and S19) when available from the genomes reported in this study and those planctomycetal genomes available from the NCBI and IMG/M[85] in April 2017 (165 genomes in total). All ribosomal proteins were obtained from the Prokka[76] annotation (see above) and the sequences of each protein were aligned using MUSCLE v.3.8.31 (ref.[87]) and reduced with Gblocks v0.91b[88]. The sequences were then concatenated. The phylogenetic analysis was done using either an ML approach (1,000 bootstraps, model of amino acid substitution: PROTGAMMAIWAG) or BI (rate variation model: invgamma, amino acid model: fixed(wag), single chain per analysis, number of generations: 375,000)[89,90]. Ribosomal protein sequences from two verrucomicrobial strains served as the outgroup. The resulting trees are in Newick format in Supplementary Data 14 and 15.

Gene-content analysis

The analysis of the gene content was performed with all of the genomes reported in this study and all planctomycetal genomes available from the NCBI and IMG/M85 in April 2017 with a contamination rate <10% (ref.[72]). The completeness[72] of the genomes was >90% (208 genomes in total). Orthologous groups for all genomes were determined based on a best-bidirectional BLAST approach using Proteinortho[86]. The e-value threshold was set to 1×10^{-8}, the identity threshold to 30% and the query coverage threshold to 70%. Singleton genes were removed from the analysis. The results were translated to binary characters, based on the absence or presence of genes in the orthologous groups. The Jaccard distance was determined using the dist function of the R package 'stats'[91]. Neighbour-joining phylogeny with 1,000 bootstrap values was calculated with the root function of the R package 'ape'[92]. Two verrucomicrobial strains served as the outgroup. The resulting trees are in Newick format in Supplementary Data 16.

Amino acid identity

The analysis of the amino acid identity was performed with the same genomes as in the 'Gene-content analysis' section. The amino acid identity was calculated with the aai.rb script of the enveomics collection[93]. The resulting values were transformed into a distance matrix and a neighbour-joining tree without bootstrap values was calculated with the nj function of the R package 'ape'[92]. Two verrucomicrobial strains served as the outgroup. The resulting trees are in Newick format in Supplementary Data 17.

Supertree analysis

All 16 gained trees (1.7.1 to 1.7.6) were concatenated and served as input for a supertree analysis with PhySIC_IST[94]. PhySIC_IST was run with a bootstrap threshold of zero and a correction threshold of one (veto method), and with a bootstrap threshold of 70 and a correction threshold of 0.6 (voting method). The resulting trees are in Newick format in Supplementary Data 18 and 19. The results were manually transferred to the MLSA tree 'c90cov50ML' (compare Supplementary Data 2) and are depicted in Supplementary Fig. 1.

Environmental occurrence of isolates and their close relatives

The 16S rRNA was analyzed using IMNGS[95] with a 99% sequence-identity threshold invoked. Briefly, the NCBI Sequence Read Archive (March 2019) was searched for the occurrence of matching operational taxonomic units and, if present, the metadata of the respective study were provided. Datasets with >0.02% hits against planctomycetes (total) were included in the further analysis. If >0.005% of the hits of a dataset were for a specific strain, the dataset was counted with '1' to design the pie charts in Fig. 1.

Pan and core genome analysis

For the detailed analysis of the planctomycetal core and pan genome (Supplementary Figs. 4 and 5), data were reused from the pan-genome matrix generated earlier (see the 'MLSA' section in the Methods). To categorize and compare the pan genome of individual planctomycetal clades, the classical categories 'strict core', 'soft core', 'shell' and 'cloud' coined by Kaas *et al.*[96] and Koonin and Wolf[97] were used. To take into account the still

relatively small size of some of the clades, some categories were slightly modified as follows. The strict core comprised orthologous clusters present in all genomes and the soft core comprised clusters found in more than 90% of the genomes. The shell contained moderately common genes present in more than 75% (shell ≥75%) or more than 50% (shell ≥50%) of the orthologous gene clusters. The cloud comprised all those genes present in less than 50% (cloud <50%), less than 25% (cloud <25%) or only one (singletons) of the genomes (Supplementary Table 3). For Supplementary Fig. 4, each subcategory was further divided into fractions – based on ortholog abundance in other planctomycetal clades ('Not in other clades', 'Shell/cloud in other clades', 'Core in other clades')—allowing, for example, the identification of characteristic clade-specific gene fractions (predominant in one clade but uncommon or absent in other clades). The 150 genomes included in the calculation of the pan and core genome have been plotted in Supplementary Fig. 5. One exception was made for the calculation of the strict core genome (purple): to compensate for genes missing due to poor assembly qualities of the genomes, only those 115 genomes with <90 scaffolds were included in the analysis. For a list of all strict and soft core genome genes, COG categorization[98] and a list of a manually revised annotation[99,100] refer to Supplementary Table 4.

Presence of genes canonically involved in cell division, cytoskeleton formation and peptidoglycan biosynthesis

A detailed analysis of the absence or presence of genes that are known to play an important part in cell division and peptidoglycan synthesis was performed. The examined bacterial genomes were either sequenced in this study or were publicly available in April 2017; they all had to have an estimated completeness[72] >90% and an estimated contamination[72] level <5% (150 in total). All externally acquired genomes were re-annotated using Prokka v1.11 (ref.[76]) to ensure the comparability of the results. The targets for the analysis were chosen from the review articles by Ingerson-Mahar and Gitai[101], Graumann[102] and Du and Lutkenhaus[20] as well as the original research article by Jogler and colleagues[22]. The Pfam[103] family and the Conserved Domain Database[100] identifiers were assembled for all of the genes in question. Following the download of a Pfam alignment, the profile Hidden Markov Models (HMM) was computed using hmmbuild[104]. The profile HMM was then used to inspect all genomes with hmmsearch[104]. In addition, the position-specific scoring matrix linked to the Conserved Domain Database entry was used to perform a reverse position-specific BLAST[105] search with all of the genes of all genomes as the query. In some cases, a BLASTp[106] analysis with genes from *P. limnophila*, *E. coli* or *B. subtilis* was performed. During the subsequent manual inspection process, the necessary alignments were done using MUSCLE v.3.8.31 (ref.[87]) and trees for visualization were calculated with FastTree 2.1.9 (ref.[83]). The results of this analysis can be found in Supplementary Table 5 and Supplementary Figs. 7–10. More detailed specification (and certain results) for every examined gene can be found in Supplementary Table 12.

Depletion of genes canonically involved in cell division

The plasmids used for gene deletion in a double event of homologous recombination were derived from the pEX18Tc vector[107], which is a plasmid containing a tetracycline resistance gene. To construct knockout plasmids for *ftsI*, *ftsW* and *ftsK*, fragments containing 1,000–

1,300 bp sequences upstream and downstream of the target genes were amplified by PCR from the genomic DNA of *P. limnophila* using the primers specified in Supplementary Table 13.

The digested upstream and downstream fragments were then cloned into pEX18Tc. Finally, the kanamycin-resistance gene from the pUTminiTn5km plasmid[108] was amplified using the primers Km BamHI fwd and Km BamHI rv, cut with BamHI and subsequently cloned as a BamHI fragment between the two flanking regions.

Transformation of P. limnophila was performed by electroporation as described before[109]. Fresh electrocompetent cells from 400 ml of a culture at an optical density (600 nm) of 0.4 in modified PYGV were prepared. The cells were washed twice with ice-cold double-distilled sterile water (100 and 50 ml) and once with 2 ml ice-cold 10% glycerol. The pellet was then resuspended in 400 µl ice-cold 10% glycerol and aliquots of 100 µl were dispensed into 0.1-mm gapped electroporation cuvettes along with approximately 1 µg circular DNA and 1 µl Type-One restriction inhibitor (Epicentre). Electroporation was performed using a Bio-Rad Micropulser (Ec3 pulse; voltage, 3.0 kV). The electroporated cells were immediately recovered in 1 ml cold medium and incubated at 28 °C for 2 h with shaking. The cells were then plated onto selective plates supplemented with 50 µg ml^{-1} kanamycin and incubated at 28 °C until colony formation after 7–9 days. The colonies were segregated onto fresh selection plates and verified by sequencing.

Analysis of potential secondary metabolite-producing BGCs

All of the genomes gained during the course of this study and genomes that were publicly available at the NCBI and IMG/M in May 2017 were included in the analysis. We excluded genomes that were derived from metagenome binning and have a contamination value[72] >5%. In the first step, the genomes were screened for the presence of BGCs using AntiSMASH 3.0 (ref.[48]; parameters: –limit −1 –disable-BiosynML –clusterblast –knownclusterblast –smcogs). In the next step, the identified BGCs were clustered into groups of related genomic content by employing BiG-SCAPE47 (parameters: –include_singletons –mix). To visualize the identified BGCs (see also Supplementary Figs. 33–44) as well as the links between the clusters they form and the phylogenetic group of the source organisms, a circular plot was created with Circos110. Cytoscape v3.6.0 (ref.[111]) was used to display the gained similarity network with a raw distance threshold smaller than 0.8 (parameters: prefuse force directed layout, no distance; Supplementary Figs. 34–44). In addition, the described BGCs of the MIBiG46 repository as well as all BGCs from the analyzed planctomycetal genomes were clustered using BiG-SCAPE[47] (Supplementary Fig. 45).

Analysis of giant genes

In the first step, one representative genomic nucleotide sequence was downloaded from the NCBI ftp server in May 2017 with the help of pyani v0.2.7 (ref.[112]) for all bacterial species with available genomes. The genomes had to be closed and had to contain fewer than 20 gaps (of stretches of 20 Ns). To ensure the comparability of the results, a new open reading frame calling with Prodigal V2.6.3 (ref.[113]) was performed for all genomes. The number of

genes that were >5,000 bp was calculated from the acquired dataset. In addition, an RNAmmer 1.2 (ref.[114]) analysis was performed to extract the 16S rRNA sequences of all genomes. A phylogenetic tree was built with FastTree 2.1.8 (ref.[83]) with the found sequences (Supplementary Fig. 46). The same steps were applied to all planctomycetal genomes that matched the mentioned criteria regarding genome status and gap number.

Signal transduction systems analysis

The examined bacterial genomes were either sequenced in this study or were publicly available in April 2017; they all had to have an estimated completeness[72] >90% and an estimated contamination[72] level <5% (150 in total). To ensure the comparability of the results, all externally acquired genomes were re-annotated with Prokka v1.11 (ref.[76]). The numbers of signaling proteins were determined essentially as described previously from the results of clade-specific Position-Specific Iterated BLAST searches, followed by the case-by-case manual evaluation of borderline hits[36,115].

ECF σ factor analysis

The analysis was performed with the same genomes as in the ‘Signal transduction systems analysis’. The HMMER software[116] was used to scan the translated amino acid sequences for the presence of the relevant Pfam version 29.0 (ref.[103]) conserved protein domains (PF03979: Sigma70_r1_1; PF00140: Sigma70_r1_2; PF04546: Sigma70_ner; PF04542: Sigma70_r2; PF04539: Sigma70_r3; PF04545: Sigma70_r4; PF08281: Sigma70_r4_2: PF07638: Sigma70_ECF) using a reporting threshold e-value of 10.0. All amino acid sequences containing conserved protein domains Sigma70_r2 and Sigma70_r4 or Sigma70_r4_2 were selected, whereas those also containing domains Sigma70_ r1_1, Sigma70_r1_2, Sigma70_ner or Sigma70_r3 were excluded. The HMMER software[116] was again used to scan the remaining sequences against a database of ECF group HMM models. A database was built from Clustal Omega alignments[117] of all of the sequences used to define each of the published 94 ECF groups[22,38,39,118]. All of the HMM models were tested against all sequences and the trusted scores (that is, the lowest score generated by a true member of the corresponding ECF group) were used as inclusion thresholds. All of the ECF sequences were aligned using Clustal Omega[117], as implemented in CLC Main Workbench 7.9.1 (Qiagen Bioinformatics), and groups were defined based on the simultaneously generated phylogenetic tree, build with the neighbour-joining grouping method and without distance corrections. All ECF amino acid sequences were analyzed for the presence of additional conserved protein domains using Pfam version 29.0 and the domain-specific gathering cutoffs, as implemented in CLC Main Workbench 7.9.1 (Qiagen Bioinformatics). In addition, those sequences were submitted to the TMHMM server version 2.0 (ref.[119]) for the prediction of transmembrane helices. Genomic context conservation was analyzed for each ECF group. Members of each group were annotated back into the genome and the ECF-encoding gene sequences together with 5,000 bp up- and downstream were extracted and aligned using Clustal Omega[117], as implemented in CLC Main Workbench 7.9.1 (Qiagen Bioinformatics). The putative functions of the conserved neighbouring gene products were inferred based on their initial annotation as well as by the presence of Pfam conserved protein domains[103] and transmembrane helices as predicted by the implementation of TMHMM[119] in CLC Main Workbench 7.9.1 (Qiagen Bioinformatics).

Manual curation of the ECF grouping was performed to merge groups sharing the same characteristics and divide those that contained ECF σ factors with different characteristics. Moreover, to identify missed members of all ECF groups, the sequences left ungrouped by this approached were analyzed by a BLAST[106] search against a local database build from the amino acid sequences of the ones that were grouped; the ECF σ factors with high similarity to members of a given group were assigned to that group. Multiple sequence alignments of the ECF amino acid sequences of each of the refined groups were performed with Clustal Omega[117]; based on such alignments, a pairwise comparison of the Jukes–Cantor distance among all sequences was generated and the outliers were removed from the groups.

Data and strain availability

Genome and 16S rRNA gene sequences have been submitted to the NCBI and are available under the accession numbers provided n Supplementary Table 1, and are covered by the NCBI BioProject numbers PRJNA485700, PRJNA522732, PRJNA357569, PRJNA522774 and PRJNA362528. The availability of the axenic cultures reported in this study is also stated in detail in Supplementary Table 1.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgments

We thank Laura van Niftrik for providing bacterial biomass from anaerobic lab-scale bioreactors. We highly appreciate the help of Christine Wiegand and Amadeus Scharmach in naming the novel strains appropriately according to community standards. The GHOSTDABS project kindly provided the left-most image in the upper panel of in Figure. 1. We further thank Jörn Piel for scientific discussion and Cathrin Spröer for help with sequencing the planctomycetal strains. This work was kindly funded by the Deutsche Forschungsgemeinschaft DFG JO 893/4-1 and the Volkswagen foundation (Experiment! Nr. 89256). MYG was supported by the NIH Intramural Research Program at the U.S. National Library of Medicine.

References

1. Wiegand S, Jogler M & Jogler C On the maverick Planctomycetes. FEMS Microbiol. Rev 42, 739–760 (2018). [PubMed: 30052954]
2. Wagner M & Horn M The *Planctomycetes*, *Verrucomicrobia*, *Chlamydiae* and sister phyla comprise a superphylum with biotechnological and medical relevance. Curr. Opin. Biotechnol 17, 241–249 (2006). [PubMed: 16704931]
3. Peeters SH & van Niftrik L Trending topics and open questions in anaerobic ammonium oxidation. Curr. Opin. Chem. Biol 49, 45–52 (2018). [PubMed: 30308437]
4. Jeske O et al. Developing techniques for the utilization of Planctomycetes as producers of bioactive molecules. Front. Microbiol 7, 1242 (2016). [PubMed: 27594849]
5. van Teeseling MCF et al. Anammox Planctomycetes have a peptidoglycan cell wall. Nat. Commun 6, 6878 (2015). [PubMed: 25962786]
6. Jeske O et al. Planctomycetes do possess a peptidoglycan cell wall. Nat. Commun 6, 7116 (2015). [PubMed: 25964217]
7. Boedeker C et al. Determining the bacterial cell biology of Planctomycetes. Nat. Commun 8, 14853 (2017). [PubMed: 28393831]
8. Devos DP Re-interpretation of the evidence for the PVC cell plan supports a Gram-negative origin. Antonie Van Leeuwenhoek 105, 271–274 (2014). [PubMed: 24292377]

9. Bondoso J et al. *Rhodopirellula lusitana* sp. nov. and *Rhodopirellula rubra* sp. nov., isolated from the surface of macroalgae. Syst. Appl. Microbiol 37, 157–164 (2014). [PubMed: 24631661]
10. Hirsch P & Müller M *Planctomyces limnophilus* sp. nov., a stalked and budding bacterium from freshwater. Syst. Appl. Microbiol 6, 276–280 (1985).
11. Kulichevskaya IS et al. *Zavarzinella formosa* gen. nov., sp. nov., a novel stalked, *Gemmata*-like planctomycete from a Siberian peat bog. Int. J. Syst. Evol. Microbiol 59, 357–364 (2009). [PubMed: 19196778]
12. Jeske O, Jogler M, Petersen J, Sikorski J & Jogler C From genome mining to phenotypic microarrays: *Planctomycetes* as source for novel bioactive molecules. Antonie Van Leeuwenhoek 104, 551–567 (2013). [PubMed: 23982431]
13. O'Malley MA The nineteenth century roots of 'everything is everywhere'. Nat. Rev. Microbiol 5, 647–651 (2007). [PubMed: 17603517]
14. Storesund JE, Lanzen A, Garcia-Moyano A, Reysenbach AL & Øvreås L Diversity patterns and isolation of Planctomycetes associated with metalliferous deposits from hydrothermal vent fields along the Valu Fa Ridge (SW Pacific). Antonie Van Leeuwenhoek 111, 841–858 (2018). [PubMed: 29423768]
15. Storesund JE & Øvreås L Diversity of Planctomycetes in iron-hydroxide deposits from the arctic mid ocean ridge (AMOR) and description of Bythopirellula goksoyri gen. nov., sp. nov., a novel Planctomycete from deep sea iron-hydroxide deposits. Antonie Van Leeuwenhoek 104, 569–584 (2013). [PubMed: 24018702]
16. Yarza P et al. Uniting the classification of cultured and uncultured bacteria and archaea using 16S rRNA gene sequences. Nat. Rev. Microbiol 12, 635–645 (2014). [PubMed: 25118885]
17. Galperin MY, Kristensen DM, Makarova KS, Wolf YI & Koonin EV Microbial genome analysis: the COG approach. Brief. Bioinform 20, 1063–1070 (2017).
18. Galperin MY, Makarova KS, Wolf YI & Koonin EV Expanded microbial genome coverage and improved protein family annotation in the COG database. Nucleic Acids Res. 43, D261–D269 (2015). [PubMed: 25428365]
19. Jogler C, Glöckner FO & Kolter R Characterization of *Planctomyces limnophilus* and development of genetic tools for its manipulation establish it as a model species for the phylum *Planctomycetes*. Appl. Environ. Microbiol 77, 5826–5829 (2011). [PubMed: 21724885]
20. Du S & Lutkenhaus J Assembly and activation of the *Escherichia coli* divisome. Mol. Microbiol 105, 177–187 (2017). [PubMed: 28419603]
21. Wagstaff J & Löwe J Prokaryotic cytoskeletons: protein filaments organizing small cells. Nat. Rev. Microbiol 16, 187–201 (2018). [PubMed: 29355854]
22. Jogler C et al. Identification of proteins likely to be involved in morphogenesis, cell division, and signal transduction in *Planctomycetes* by comparative genomics. J. Bacteriol 194, 6419–6430 (2012). [PubMed: 23002222]
23. Pilhofer M et al. Characterization and evolution of cell division and cell wall synthesis genes in the bacterial phyla Verrucomicrobia, Lentisphaerae, Chlamydiae, and Planctomycetes and phylogenetic comparison with rRNA genes. J. Bacteriol 190, 3192–3202 (2008). [PubMed: 18310338]
24. Rued BE et al. Structure of the large extracellular loop of FtsX and its interaction with the essential peptidoglycan hydrolase PcsB in *Streptococcus pneumoniae*. mBio 10, e02622–18 (2019).
25. Fuerst JA & Sagulenko E Beyond the bacterium: Planctomycetes challenge our concepts of microbial structure and function. Nat. Rev. Microbiol 9, 403–413 (2011). [PubMed: 21572457]
26. van Teeseling MCF, de Pedro MA & Cava F Determinants of bacterial morphology: from fundamentals to possibilities for antimicrobial targeting. Front. Microbiol 8, 1264 (2017). [PubMed: 28740487]
27. Shi H, Bratton BP, Gitai Z & Huang KC How to build a bacterial cell: MreB as the foreman of *E. coli* construction. Cell 172, 1294–1305 (2018). [PubMed: 29522748]
28. Hussain S et al. MreB filaments align along greatest principal membrane curvature to orient cell wall synthesis. eLife 7, e32471 (2018). [PubMed: 29469806]
29. Waidner B et al. A novel system of cytoskeletal elements in the human pathogen *Helicobacter pylori*. PLoS Pathog. 5, e1000669 (2009). [PubMed: 19936218]

30. Jacquier N, Viollier PH & Greub G The role of peptidoglycan in chlamydial cell division: towards resolving the chlamydial anomaly. FEMS Microbiol. Rev 39, 262–275 (2015). [PubMed: 25670734]
31. Ouellette SP, Karimova G, Subtil A & Ladant D *Chlamydia* co-opts the rod shape-determining proteins MreB and Pbp2 for cell division. Mol. Microbiol 85, 164–178 (2012). [PubMed: 22624979]
32. Jacquier N, Frandi A, Pillonel T, Viollier PH & Greub G Cell wall precursors are required to organize the chlamydial division septum. Nat. Commun 5, 3578 (2014). [PubMed: 24709914]
33. Galperin MY What bacteria want. Environ. Microbiol 20, 4221–4229 (2018). [PubMed: 30187651]
34. Jacob-Dubuisson F, Mechaly A, Betton J-M & Antoine R Structural insights into the signalling mechanisms of two-component systems. Nat. Rev. Microbiol 16, 585–593 (2018). [PubMed: 30008469]
35. Campagne S, Allain FH & Vorholt JA Extra cytoplasmic function sigma factors, recent structural insights into promoter recognition and regulation. Curr. Opin. Struct. Biol 30, 71–78 (2015). [PubMed: 25678040]
36. Galperin MY, Makarova KS, Wolf YI & Koonin EV Phyletic distribution and lineage-specific domain architectures of archaeal two- component signal transduction systems. J. Bacteriol 200, e00681–17 (2018). [PubMed: 29263101]
37. Mascher T Signaling diversity and evolution of extracytoplasmic function (ECF) sigma factors. Curr. Opin. Microbiol 16, 148–155 (2013). [PubMed: 23466210]
38. Staron A et al. The third pillar of bacterial signal transduction: classification of the extracytoplasmic function (ECF) sigma factor protein family. Mol. Microbiol 74, 557–581 (2009). [PubMed: 19737356]
39. Huang X, Pinto D, Fritz G & Mascher T Environmental sensing in Actinobacteria: a comprehensive survey on the signaling capacity of this phylum. J. Bacteriol 197, 2517–2535 (2015). [PubMed: 25986905]
40. Pinto D & Mascher T in Stress and Environmental Regulation of Gene Expression and Adaptation in Bacteria (ed. de Bruijn FJ) Ch. 2.6 (Wiley–Blackwell, 2016).
41. Bayer-Santos E et al. *Xanthomonas citri* T6SS mediates resistance to *Dictyostelium* predation and is regulated by an ECF σ factor and cognate Ser/Thr kinase. Environ. Microbiol 20, 1562–1575 (2018). [PubMed: 29488354]
42. Castro AN, Lewerke LT, Hastle JL & Ellermeier CD Signal peptidase is necessary and sufficient for site 1 cleavage of RsiV in *Bacillus subtilis* in response to lysozyme. J. Bacteriol 200, e00663–17 (2018). [PubMed: 29358498]
43. Tracanna V, de Jong A, Medema MH & Kuipers OP Mining prokaryotes for antimicrobial compounds: from diversity to function. FEMS Microbiol. Rev 41, 417–442 (2017). [PubMed: 28402441]
44. Calisto R et al. Anticancer activity in planctomycetes. Front. Mar. Sci 5, 499 (2019).
45. Cimermancic P et al. Insights into secondary metabolism from a global analysis of prokaryotic biosynthetic gene clusters. Cell 158, 412–421 (2014). [PubMed: 25036635]
46. Medema MH et al. Minimum information about a biosynthetic gene cluster. Nat. Chem. Biol 11, 625–631 (2015). [PubMed: 26284661]
47. Navarro-Muñoz J et al. A computational framework for systematic exploration of biosynthetic diversity from large-scale genomic data. Preprint bioRxiv https://www.biorxiv.org/content/10.1101/445270v1 (2018).
48. Weber T et al. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res. 43, W237–W243 (2015). [PubMed: 25948579]
49. Kohn T et al. *Fuerstia marisgermanicae* gen. nov., sp. nov., an unusual member of the phylum Planctomycetes from the German Wadden Sea. Front. Microbiol 7, 2079 (2016). [PubMed: 28066393]
50. Oberbeckmann S, Kreikemeyer B & Labrenz M Environmental Factors support the formation of specific bacterial assemblages on microplastics. Front. Microbiol 8, 2709 (2018). [PubMed: 29403454]

51. Sipkema D et al. Multiple approaches to enhance the cultivability of bacteria associated with the marine sponge *Haliclona* (*gellius*) sp. Appl. Environ. Microbiol 77, 2130–2140 (2011). [PubMed: 21296954]
52. Rast P et al. Three novel species with peptidoglycan cell walls form the new genus *Lacunisphaera* gen. nov. in the family Opitutaceae of the verrucomicrobial subdivision 4. Front. Microbiol 8, 202 (2017). [PubMed: 28243229]
53. Schlesner H 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 (1994).
54. Lage OM & Bondoso J Bringing *Planctomycetes* into pure culture. Front. Microbiol 3, 405 (2012). [PubMed: 23335915]
55. Pascual J et al. *Roseisolibacter agri* gen. nov., sp. nov., a novel slow-growing member of the under-represented phylum *Gemmatimonadetes*. Int. J. Syst. Evol. Microbiol 68, 1028–1036 (2018). [PubMed: 29458671]
56. van Kessel MAHJ et al. Complete nitrification by a single microorganism. Nature 528, 555–559 (2015). [PubMed: 26610025]
57. Bolger AM, Lohse M & Usadel B Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics 30, 2114–2120 (2014). [PubMed: 24695404]
58. O'Connell J et al. NxTrim: optimized trimming of Illumina mate pair reads. Bioinformatics 31, 2035–2037 (2015). [PubMed: 25661542]
59. Wingett SW & Andrews S FastQ screen: a tool for multi-genome mapping and quality control. F1000Res. 7, 1338 (2018). [PubMed: 30254741]
60. Schmieder R & Edwards R Quality control and preprocessing of metagenomic datasets. Bioinformatics 27, 863–864 (2011). [PubMed: 21278185]
61. Magoc T & Salzberg SL FLASH: fast length adjustment of short reads to improve genome assemblies. Bioinformatics 27, 2957–2963 (2011). [PubMed: 21903629]
62. Bankevich A et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol 19, 455–477 (2012). [PubMed: 22506599]
63. Zerbino DR & Birney E Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 18, 821–829 (2008). [PubMed: 18349386]
64. Li H & Durbin R Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25, 1754–1760 (2009). [PubMed: 19451168]
65. Koboldt DC et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. Genome Res. 22, 568–576 (2012). [PubMed: 22300766]
66. McKenna A et al. The genome analysis toolkit: a mapreduce framework for analyzing next-generation DNA sequencing data. Genome Res. 20, 1297–1303 (2010). [PubMed: 20644199]
67. Antipov D, Korobeynikov A, McLean JS & Pevzner PA hybridSPAdes: an algorithm for hybrid assembly of short and long reads. Bioinformatics 32, 1009–1015 (2016). [PubMed: 26589280]
68. Boetzer M, Henkel CV, Jansen HJ, Butler D & Pirovano W Scaffolding pre-assembled contigs using SSPACE. Bioinformatics 27, 578–579 (2011). [PubMed: 21149342]
69. Paulino D et al. Sealer: a scalable gap-closing application for finishing draft genomes. BMC Bioinform. 16, 230 (2015).
70. Bosi E et al. MeDuSa: a multi-draft based scaffolder. Bioinformatics 31, 2443–2451 (2015). [PubMed: 25810435]
71. Green MR & Sambrook J Molecular Cloning: A Laboratory Manual. 4th edn (Cold Spring Harbor Laboratory Press, 2012).
72. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P & Tyson GW CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 25, 1043–1055 (2015). [PubMed: 25977477]
73. Gurevich A, Saveliev V, Vyahhi N & Tesler G QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075 (2013). [PubMed: 23422339]
74. Robinson JT et al. Integrative genomics viewer. Nat. Biotechnol 29, 24–26 (2011). [PubMed: 21221095]

75. Langmead B & Salzberg SL Fast gapped-read alignment with bowtie 2. Nat. Methods 9, 357–359 (2012). [PubMed: 22388286]
76. Seemann T Prokka: rapid prokaryotic genome annotation. Bioinformatics 30, 2068–2069 (2014). [PubMed: 24642063]
77. Michaelis W et al. Microbial reefs in the black sea fueled by anaerobic oxidation of methane. Science 297, 1013–1015 (2002). [PubMed: 12169733]
78. Meyerdierks A et al. Insights into the genomes of archaea mediating the anaerobic oxidation of methane. Environ. Microbiol 7, 1937–1951 (2005). [PubMed: 16309392]
79. Meyerdierks A et al. Metagenome and mRNA expression analyses of anaerobic methanotrophic archaea of the ANME-1 group. Environ. Microbiol 12, 422–439 (2010). [PubMed: 19878267]
80. Muyzer G, Teske A, Wirsen CO & Jannasch HW Phylogenetic relationships of Thiomicrospira species and their identification in deep-sea hydrothermal vent samples by denaturing gradient gel electrophoresis of 16S rDNA fragments. Arch. Microbiol 164, 165–172 (1995). [PubMed: 7545384]
81. Yilmaz P et al. The SILVA and “All-species living tree project (LTP)” taxonomic frameworks. Nucleic Acids Res. 42, D643–D648 (2014). [PubMed: 24293649]
82. Pruesse E, Peplies J & Glöckner FO SINA: accurate high-throughput multiple sequence alignment of ribosomal RNA genes. Bioinformatics 28, 1823–1829 (2012). [PubMed: 22556368]
83. Price MN, Dehal PS & Arkin AP FastTree 2—approximately maximum-likelihood trees for large alignments. PloS ONE 5, e9490 (2010). [PubMed: 20224823]
84. Letunic I & Bork P Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 1, gkz239 (2019).
85. Markowitz VM et al. IMG: the integrated microbial genomes database and comparative analysis system. Nucleic Acids Res. 40, D115–D122 (2012). [PubMed: 22194640]
86. Lechner M et al. Proteinortho: detection of (co-)orthologs in large-scale analysis. BMC Bioinform. 12, 124 (2011).
87. Edgar RC MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797 (2004). [PubMed: 15034147]
88. Castresana J Selection of conserved blocks from multiple alignments for their use in phylogenetic analysis. Mol. Biol. Evol 17, 540–552 (2000). [PubMed: 10742046]
89. Stamatakis A RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30, 1312–1313 (2014). [PubMed: 24451623]
90. Ronquist F & Huelsenbeck JP MrBayes 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19, 1572–1574 (2003). [PubMed: 12912839]
91. R: a language and environment for statistical computing (R Foundation for Statistical Computing, 2017).
92. Paradis E, Claude J & Strimmer K APE: analyses of phylogenetics and evolution in R language. Bioinformatics 20, 289–290 (2004). [PubMed: 14734327]
93. Rodriguez-R LM & Konstantinidis KT The enveomics collection: a toolbox for specialized analyses of microbial genomes and metagenomes. Preprint at PeerJ Prepr. 4, e1900v1901 (2016).
94. Scornavacca C, Berry V, Lefort V, Douzery EJ & Ranwez V PhySIC_IST: cleaning source trees to infer more informative supertrees. BMC Bioinform. 9, 413 (2008).
95. Lagkouvardos I et al. IMNGS: a comprehensive open resource of processed 16S rRNA microbial profiles for ecology and diversity studies. Sci. Rep 6, 33721 (2016). [PubMed: 27659943]
96. Kaas RS, Friis C, Ussery DW & Aarestrup FM Estimating variation within the genes and inferring the phylogeny of 186 sequenced diverse Escherichia coli genomes. BMC Genom 13, 577 (2012).
97. Koonin EV & Wolf YI Genomics of bacteria and archaea: the emerging dynamic view of the prokaryotic world. Nucleic Acids Res. 36, 6688–6719 (2008). [PubMed: 18948295]
98. Huerta-Cepas J et al. eggNOG 4.5: a hierarchical orthology framework with improved functional annotations for eukaryotic, prokaryotic and viral sequences. Nucleic Acids Res. 44, D286–D293 (2016). [PubMed: 26582926]
99. Jones P et al. InterProScan 5: genome-scale protein function classification. Bioinformatics 30, 1236–1240 (2014). [PubMed: 24451626]

100. Marchler-Bauer A et al. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic Acids Res. 45, D200–D203 (2017). [PubMed: 27899674]
101. Ingerson-Mahar M & Gitai Z A growing family: the expanding universe of the bacterial cytoskeleton. FEMS Microbiol. Rev 36, 256–266 (2012). [PubMed: 22092065]
102. Graumann PL Cytoskeletal elements in bacteria. Annu. Rev. Microbiol 61, 589–618 (2007). [PubMed: 17506674]
103. Finn RD et al. The Pfam protein families database: towards a more sustainable future. Nucleic Acids Res. 44, D279–D285 (2016) [PubMed: 26673716]
104. Eddy SR Accelerated profile HMM searches. PLoS Comp. Biol 7, e1002195 (2011).
105. Marchler-Bauer A et al. CDD: a database of conserved domain alignments with links to domain three-dimensional structure. Nucleic Acids Res. 30, 281–283 (2002). [PubMed: 11752315]
106. Altschul SF, Gish W, Miller W, Myers EW & Lipman DJ Basic local alignment search tool. J. Mol. Biol 215, 403–410 (1990). [PubMed: 2231712]
107. Hoang TT, Karkhoff-Schweizer RR, Kutchma AJ & Schweizer HP A broad-host-range Flp-*FRT* recombination system for site-specific excision of chromosomally-located DNA sequences: application for isolation of unmarked Pseudomonas aeruginosa mutants. Gene 212, 77–87 (1998). [PubMed: 9661666]
108. Herrero M, de Lorenzo V & Timmis KN Transposon vectors containing non-antibiotic resistance selection markers for cloning and stable chromosomal insertion of foreign genes in Gram-negative bacteria. J. Bacteriol 172, 6557–6567 (1990). [PubMed: 2172216]
109. Rivas-Marin E, Canosa I, Santero E & Devos DP Development of genetic tools for the manipulation of the Planctomycetes. Front. Microbiol 7, 914 (2016). [PubMed: 27379046]
110. Krzywinski MI et al. Circos: an information aesthetic for comparative genomics. Genome Res. 19, 1639–1645 (2009). [PubMed: 19541911]
111. Shannon P et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res. 13, 2498–2504 (2003). [PubMed: 14597658]
112. Pritchard L, Glover RH, Humphris S, Elphinstone JG & Toth IK Genomics and taxonomy in diagnostics for food security: soft-rotting enterobacterial plant pathogens. Anal. Methods 8, 12–24 (2016).
113. Hyatt D et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 11, 119 (2010).
114. Lagesen K et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35, 3100–3108 (2007). [PubMed: 17452365]
115. Galperin MY A census of membrane-bound and intracellular signal transduction proteins in bacteria: bacterial IQ, extroverts and introverts. BMC Microbiol. 5, 35 (2005). [PubMed: 15955239]
116. Finn RD, Clements J & Eddy SR HMMER web server: interactive sequence similarity searching. Nucleic Acids Res. 39, W29–W37 (2011). [PubMed: 21593126]
117. Sievers F et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol 7, 539 (2011). [PubMed: 21988835]
118. Gomez-Santos N, Perez J, Sanchez-Sutil MC, Moraleda-Munoz A & Munoz-Dorado J CorE from *Myxococcus xanthus* is a copper-dependent RNA polymerase sigma factor. PLoS Genet. 7, e1002106 (2011). [PubMed: 21655090]
119. Sonnhammer EL, von Heijne G & Krogh A A hidden Markov model for predicting transmembrane helices in protein sequences. Proc. Int. Conf. Intell. Syst. Mol. Biol 6, 175–182 (1998). [PubMed: 9783223]

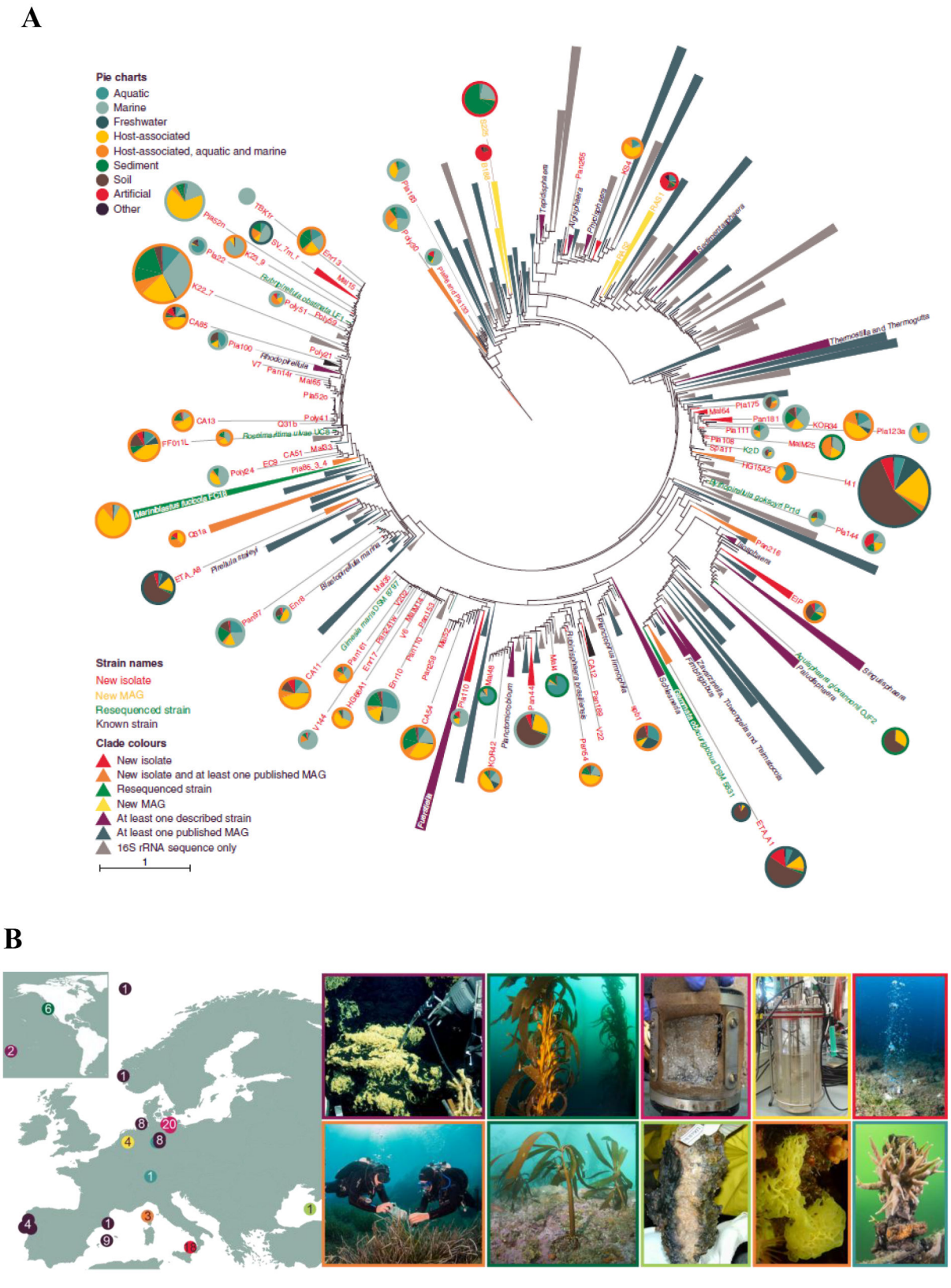

Figure 1. |. Sampling the phylum Planctomycetes.

A. Phylogeny and environmental occurrences of the planctomycetal isolates. Maximum-likelihood phylogenetic tree of 9,002 non-redundant planctomycetal 16S rRNA gene sequences, including the isolates (red) and MAGs (yellow) reported in this study as well as resequenced strains (green; Supplementary Data 1). The size of the collapsed clades corresponds to the number of sequences they comprise (between 10 and 104); clades with fewer than ten sequences have been condensed to a single line. The occurrence of the 16S rRNA genes of reported cultures in publicly available amplicon datasets from the NCBI Sequence Read Archive was analyzed with a 99% sequence-identity threshold designed to identify identical and closely related strains. The detailed distribution of the respective hits is illustrated as a corresponding pie chart for the cases where >10 hits against planctomycetes (total) were found to comprise >0.02% of a given dataset, of which >0.005% were strain specific. The number of positively screened datasets is illustrated by the respective size of the plotted pie charts (between 11 and 3,806 hits in the analyzed studies). The corresponding sectors are color-coded according to the habitat categories of the respective dataset. The outer border of each pie chart is color-coded according to the origin of the isolate itself.
B. Diversity-driven sampling. The samples for this study were obtained from a variety of aquatic locations, mostly throughout Europe, and include iron-hydroxide deposits (D); microbial mats from methane seeps in the Black Sea (K); underwater fluid discharges (gas and water) and hot lakes in the Tyrrhenian Sea around Panarea island, Italy (P); kelp surfaces at Monterey Bay, California (M); surfaces of different macroalgae and jellyfish species at Heligoland island, Germany (H); seaweed meadows and calcareous sponges of the Mediterranean Sea (C); freshwater sponges from Lake Constance (L) and Lake Salzgitter (Z), Germany; beach sediment and algae at El Arenal, Mallorca island, Spain (E); offshore seawater of Costa Brava, Spain (V); freshwater from a meromictic lake in Bergen, Norway (N); a cyanobacterial bloom in the Stadtgraben pond in Wolfenbüttel, Germany; a seawater ornamental aquarium; wood, polyethylene and polystyrene baits deposited at Warnow River estuary, Rostock, Germany and costal Baltic Sea spots, Heiligendamm, Germany (B) and hypoxic and anaerobic bioreactors (A).

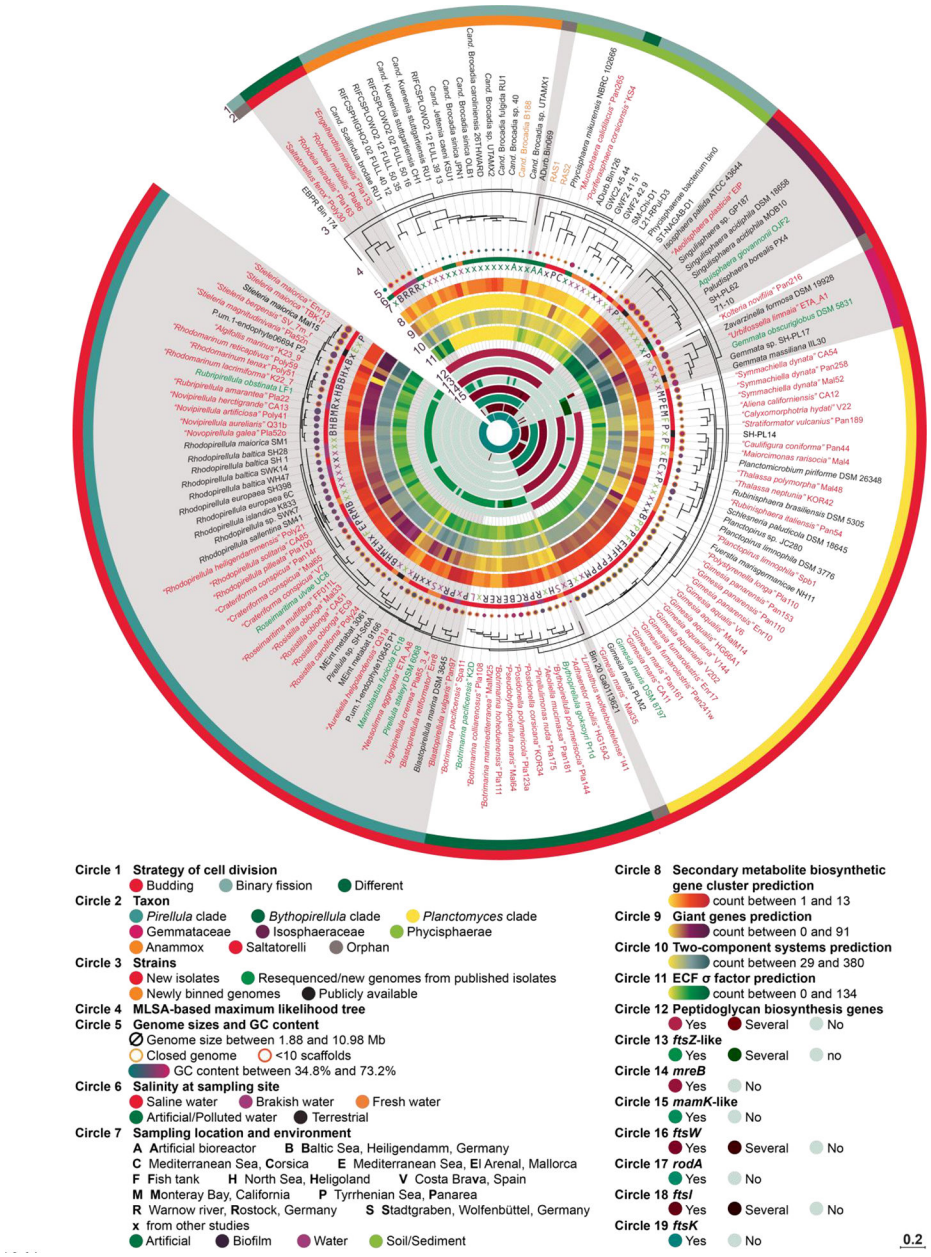

Figure 2. Current diversity of the planctomycetal phylum.

Core information assembled in this study briefly summarizing all other figures and additional data. The centrepiece of the figure is formed by a MLSA-based phylogenetic tree of all of the isolates and MAGs from this study and from the resequenced or publicly available genomes of previously described strains, enrichment cultures and metagenomic samples (a total of 150). Details on the tree as well as the rooting and branching probabilities can be found in Supplementary Fig.1, Supplementary Data 2 and Methods; details on the genomes reported in this study are given in Supplementary Table 1. The focal point is encompassed by Circles 1 and 2, which highlight the different subclades that constitute the planctomycetal phylum and the current knowledge on the distribution of different cell-division strategies throughout it. The first inner circles hold information on the sampling

sites and the environmental conditions there (Circles 6 and 7) as well as the main properties of the gathered genomes (Circle 5). They are followed by Circles 8–11, illustrating the number of identified putative secondary metabolite BGCs, giant genes, two-component systems and ECF σ factors. The innermost Circles 12–19 provide an overview of the presence or absence of the canonical genes involved in peptidoglycan biosynthesis, cytoskeleton formation and cell division.

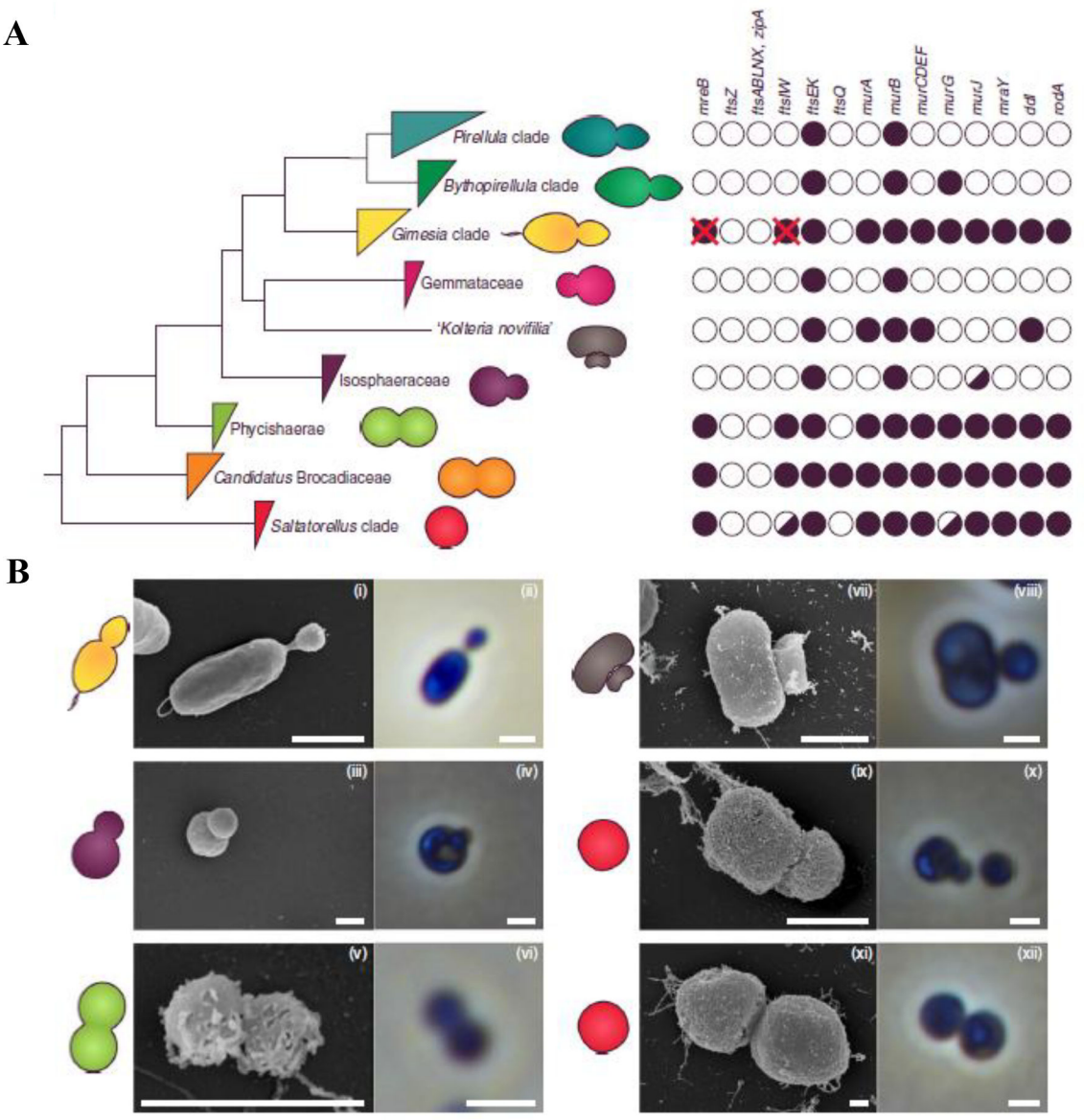

Figure 3. Planctomycetal cell division.

A. Collapsed phylogenetic tree of the planctomycetal subclades with their respective modes of cell division and the presence (filled circles) or absence (empty circles) of canonical cell-division genes (see also Supplementary Fig. 7). For an uncollapsed version of the tree please refer to Fig. 2 or Supplementary Fig. 1. The genes *mreB*, *ftsI* and *ftsW* were deleted in *P. limnophila* (marked by a red X) without observing a phenotype.

B. Most of the known planctomycetes, such as the pear-shaped *P. limnophila* ((i) and (ii)), divide by polar budding. Budding was indeed frequently observed among our isolates, not only in cells with an elongated shape but also in coccoid cells such as '*Aeolisphaera plasticia*' ElP ((iii) and (iv)). In contrast, members of the class Phycisphaerae – for example, '*Mucisphaera calidilacus*' Pan265 ((v) and (vi)) – divide by binary fission. A previously unseen mode of cell division was observed for '*Kolteria novifilia*' Pan216, which divides by lateral budding ((vii) and (viii); Supplementary Movie 1). Members of the *Saltatorellus*

clade on the other hand have the ability to perform budding ((ix) and (x); Poly30) as well as binary fission ((xi) and (xii); Poly30; Supplementary Movie 2), even in the same culture. The first and the third columns of images are scanning electron micrographs; the second and fourth columns contain phase-contrast light-microscopy images. Scale bars, 1 μm. The selected micrographs originate from at least two independent experiments; more than 100 cells were analyzed and the best representative image was chosen.

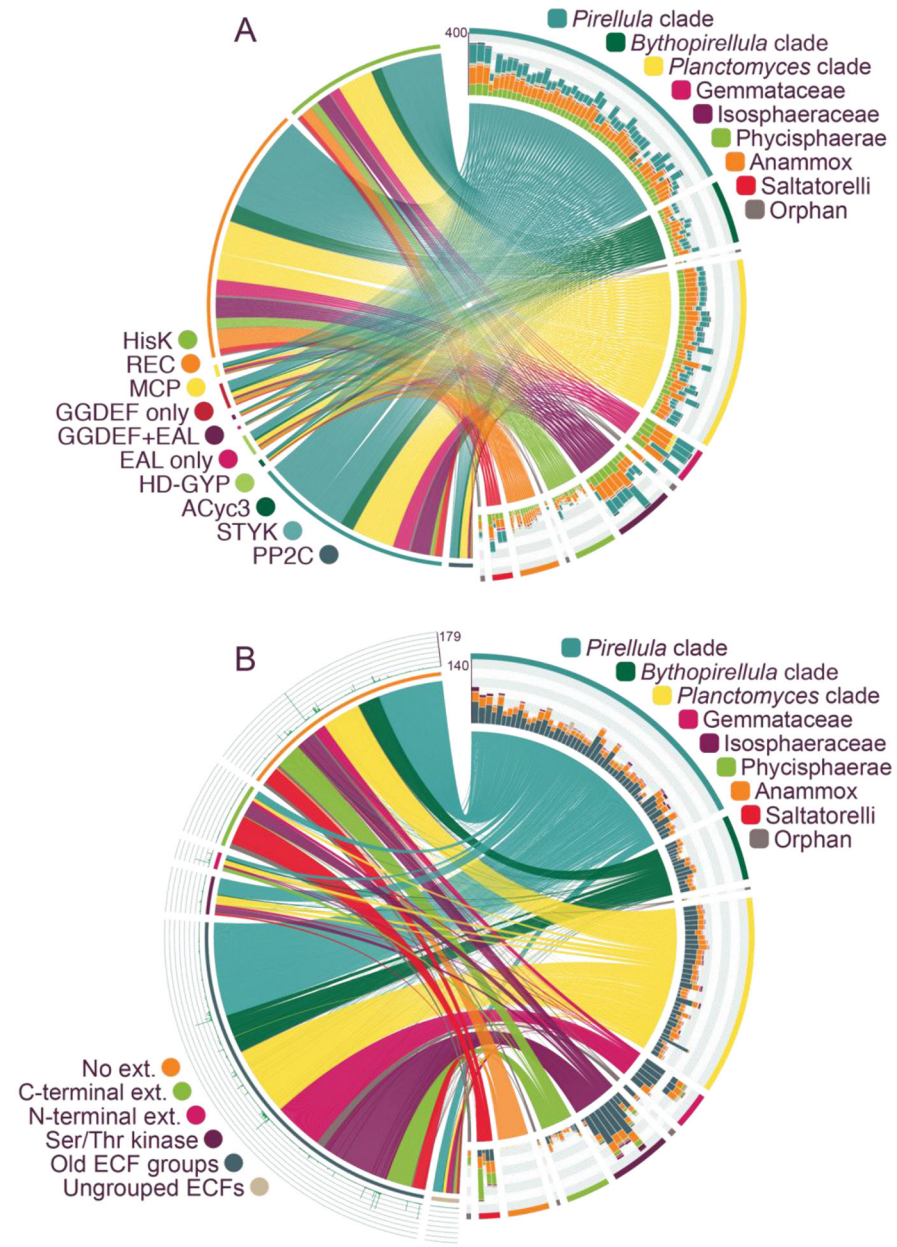

Figure 4. Signaling in Planctomycetes.

A. Signal transduction systems. A total of 26,487 components related to signal-transduction systems were identified in the 150 analyzed genomes (Supplementary Table 6). The analyzed planctomycetal genomes are depicted on the right-hand side of the circular plot. Each column of the stacked-column chart represents one strain and the different stacks of the columns outline the different signal-transduction-system components found in each genome (see the lower-left legend). The colors of the overlying arcs and the underlying links suggest the planctomycetal clade that contains the respective strain (see the upper-right legend). In addition, the links connect each strain to the category of the identified signal-transduction-system components on the left-hand side of the plot. Two-component signal-transduction systems: HisK, two-component sensor histidine kinases and REC, response regulators.

Chemotaxis: MCPs. c-di-GMP hydrolysis: GGDEF, diguanylate cyclases; EAL, EAL-domain containing c-di-GMP phosphodiesterases; HD-GyP, HD-GyP-domain containing c-di-GMP phosphodiesterases. Ser/Thr/Tyr protein kinases and phosphoprotein phosphatases: STyK, Ser/Thr/Tyr protein kinases; PP2C, PP2C-type protein phosphatases. ACyc3, class III adenylate/guanylate cyclases.

B. Extracytoplasmic function σ factors. A total of 5,966 ECF σ factors were identified in all 150 analyzed genomes (Supplementary Table 7). A subsequent similarity clustering found them to form 4,694 clusters, which are indicated on the left-hand side of the circular plot. The outer histogram marks the total number of identified ECF σ factors per cluster and the underlying curves/arcs define the groups of the ECF σ factors (see the lower-left legend). The right-hand side of the circular plot describes the analyzed planctomycetal genomes. Each column of the stacked-column chart represents one strain and the different stacks of the columns outline the ECF σ factors found in each genome (see the lower-left legend). The colors of the overlying arcs and the underlying links suggest the planctomycetal clade that encompasses the respective strain (see the upper-right legend). Finally, the links point to the ECF σ factor groups that the ECF σ factors of each strain belong to.

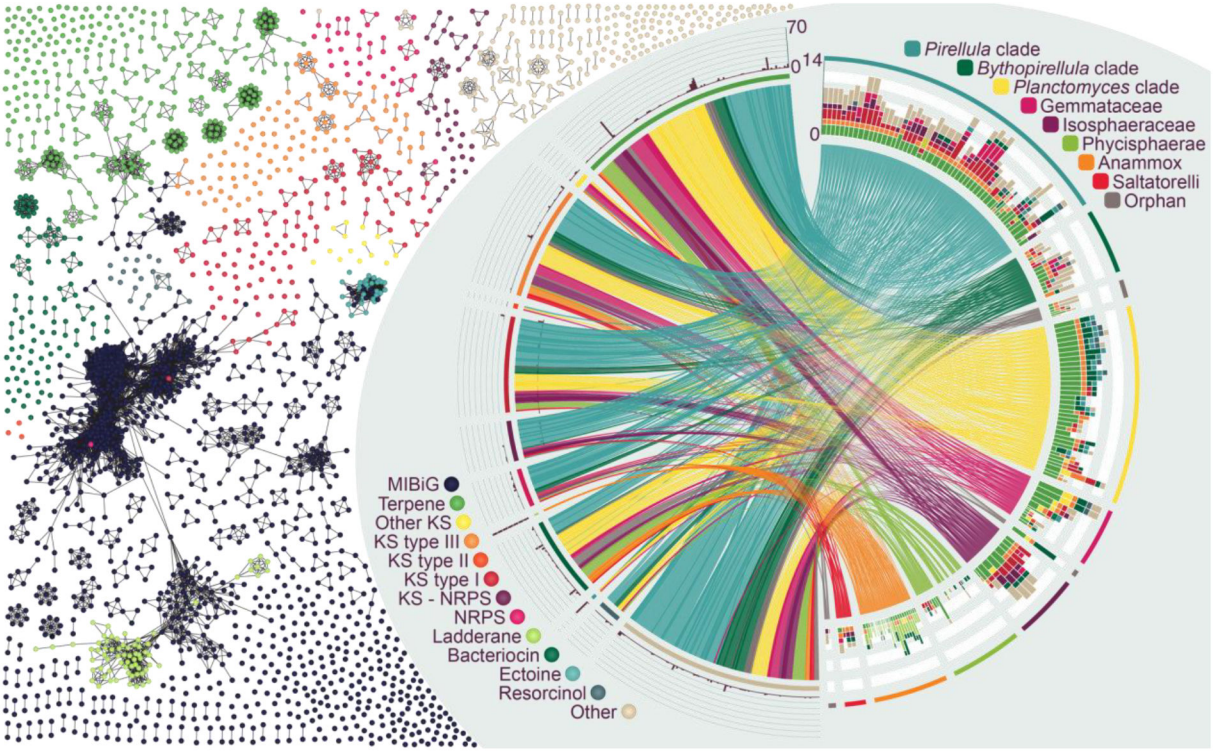

Figure 5. Secondary metabolite biosynthetic gene clusters (BGCs).

A total of 1,104 BGCs with potential involvement in the production of secondary metabolites were identified in all 196 analyzed genomes (Supplementary Table 10). Many of these BGCs fall into 162 clusters, but 359 are unique according to the applied parameters. These 521 BGC groups are provided on the left-hand side of the chart, with the histogram marking the total number of identified BGCs per group. The underlying arcs define the BGC functional category (see the lower-left legend) of the group. On the right-hand side of the chart, each column of the stacked column chart represents one strain and the different stacks of the columns sum up the BGCs found in each genome (see the lower-left legend). The colors of the overlying arcs and the connected links indicate the planctomycetal clade that encompasses the respective strain (see the upper-right legend). The links connect the hits for each strain to the corresponding BGC group. A cluster analysis of the identified BGCs with previously described BGCs is shown in Supplementary Fig. 45 and provides the relatedness of the different BGCs. NRPS, non-ribosomal peptide synthases.