

# Supplementary Material S1

## Evolution and biogeography of members of the proposed *Kiritimatiellaeota* phylum

### 1. Establishment of a novel phylum-level lineage within the PVC superphylum

To examine the robustness of the tree topology shown in Fig. 1 additional phylogenetic trees were reconstructed. The proposed separate position of the subdivision 5 was confirmed in a neighbor-joining tree based on a more comprehensive dataset containing almost all available sequences affiliated with subdivision 5 (Supplementary Fig. S1a). The obtained tree topology was also insensitive to the application of a 50% conservation filter removing all hypervariable alignment positions. In addition, trees were reconstructed based on the complete RNA polymerase beta subunit encoded by the *rpoB* gene. RpoB was chosen as alternative phylogenetic marker, because in several studies it was found that complete RpoB proteins have significantly more informative positions and are better proxies of the whole genomes of bacteria than 16S rRNA genes. Consequently, comparative sequence analyses of RpoB proteins should result in a higher phylogenetic resolution compared to 16S rRNA genes (e.g., Mollet et al., 1997; Konstantinidis and Tiedje, 2005; Spring, Scheuner, et al., 2015). The overall topology of the RpoB protein tree shown in Supplementary Fig. S1b is well supported by high bootstrap values and in good agreement with recently published trees of the PVC superphylum based on concatenated gene data sets (e.g., Rinke et al., 2013; Kamke et al., 2014). Interestingly, in the RpoB tree strain L21-Fru-AB<sup>T</sup> is separated from the remaining members of the *Verrucomicrobia* by representatives of the *Lentisphaerae* phylum, although only with insignificant bootstrap support.

### 2. Habitat preference

In a survey of the environmental distribution of 16S rRNA gene sequences the phylogenetic structure within the proposed *Kiritimatiellaeota* phylum was correlated with the environmental preferences of distinct clades. In the phylogenetic tree shown in Fig. 3b several coherent sequence clusters were identified and designated according to the first identified representative clone sequence. Note, that for reasons of clarity several minor monophyletic groups present in the SILVA database were collapsed to the clades shown in Fig. 3b, so that

the designation of clades in the shown tree do not necessarily correspond to the same set of sequences indicated in the SILVA 119 SSU Ref NR 99 database. The accession numbers, sources of isolation and respective clades of the used 16S rRNA sequences is given in Supplementary Table S1. For each clade the distribution of sequences in various habitats is illustrated as pie chart diagram.

In accordance with Fig. 1 a separation of the *Kiritimatiellaeota* phylum in at least two major lineages corresponding to class-level became apparent. Members of the class represented by strain L21-Fru-AB<sup>T</sup> seem to have a general preference for aquatic environments, whereas representatives of the other branch comprising the clades ML635J-35, WCHB1-25 and RFP12 are mainly found in the intestine of animals. A database search with the 16S rRNA gene sequence of L21-Fru-AB<sup>T</sup> as query against the NCBI nucleotide collection (nr/nt) using blastn (Altschul *et al.*, 1997) resulted in numerous hits with an identity value above 97% including the 16S rRNA sequences JN477781 and JN456175 shown in Fig. 1. All of these highly similar environmental clone sequences could be assigned to the R76-B128 clade and were retrieved from a hypersaline microbial mat located in Guerrero Negro, Mexico (Harris *et al.*, 2013), which indicates that the distribution of L21-Fru-AB<sup>T</sup> and closely related strains is restricted to hypersaline microbial mats.

Interestingly, the deeply branching phylogenetic groups of this phylum seem to comprise sequences from a broader range of environments than more recent clades, which display a pronounced habitat preference. An especially rapid expanding and fast evolving group seems to be the RFP12 clade that is comprised mainly of sequences retrieved from the intestine of vertebrates, *e.g.* the rumen of cattle. In several studies it was reported that members of the RFP12 clade were among the most abundant bacteria detected in the hindgut microflora of horses and their prevalence may even increase in response to some chronic diseases (Steelman *et al.*, 2012). In contrast, members of the deeply branching GZKB33 clade are equally distributed in wastewater, freshwater, marine and hypersaline environments and thus may still have the genetic potential to adapt to a wide range of environments. It can be speculated that representatives of this clade inhabiting freshwater sediments are most closely related with the evolutionary ancestor of this phylum, because all clades, except BS5, which is mainly found in wastewater, contain sequences retrieved from freshwater sediments.

## References

- Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, *et al.* (1997). Gapped BLAST and PSI-BLAST: A new generation of protein database search programs. *Nucleic Acids Res* **25**:3389–3402.
- Harris JK, Caporaso JG, Walker JJ, Spear JR, Gold NJ, Robertson CE, *et al.* (2013). Phylogenetic stratigraphy in the Guerrero Negro hypersaline microbial mat. *ISME J* **7**:50–60.
- Kamke J, Rinke C, Schwientek P, Mavromatis K, Ivanova N, Sczyrba A, *et al.* (2014). The candidate phylum *Poribacteria* by single-cell genomics: New insights into phylogeny, cell-compartmentation, eukaryote-like repeat proteins, and other genomic features. *PLoS One* **9**:e87353.
- Konstantinidis KT, Tiedje JM. (2005). Towards a genome-based taxonomy for prokaryotes. *J Bacteriol* **187**:6258–6264.
- Mollet C, Drancourt M, Raoult D. (1997). *rpoB* sequence analysis as a novel basis for bacterial identification. *Mol Microbiol* **26**:1005–1011.
- Rinke C, Schwientek P, Sczyrba A, Ivanova NN, Anderson IJ, Cheng J-F, *et al.* (2013). Insights into the phylogeny and coding potential of microbial dark matter. *Nature* **499**:431–437.
- Spring S, Scheuner C, Göker M, Klenk H-P. (2015). A taxonomic framework for emerging groups of ecologically important marine gammaproteobacteria based on the reconstruction of evolutionary relationships using genome-scale data. *Front Microbiol* **6**:281.
- Steelman SM, Chowdhary BP, Dowd S, Suchodolski J, Janečka JE. (2012). Pyrosequencing of 16S rRNA genes in fecal samples reveals high diversity of hindgut microflora in horses and potential links to chronic laminitis. *BMC Vet Res* **8**:231.