

A novel and diverse group of Patescibacteria from bathypelagic Lake Baikal revealed through long-read metagenomics

Jose M. Haro-Moreno^a, Pedro J. Cabello-Yeves^a, M. Pilar Garcillán-Barcia^b, Alexandra Zakharenko^c, Tamara I. Zenskaya^c, and Francisco Rodriguez-Valera^{a,*}

^aEvolutionary Genomics Group, Departamento Producción Vegetal y Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan 03550, Alicante, Spain. ^bInstituto de Biomedicina y Biotecnología de Cantabria (IBBTEC), Universidad de Cantabria-Consejo Superior de Investigaciones Científicas, Santander, Spain.

^cLimnological Institute, Siberian Branch of the Russian Academy of Sciences, Irkutsk, Russia.

*Corresponding author: frvalera@umh.es

Evolutionary Genomics Group, Departamento Producción Vegetal y Microbiología, Universidad Miguel Hernández, Apartado 18, San Juan de Alicante, 03550 Alicante, Spain.

Phone +34-965919313, Fax +34-965 919457

SUPPORTING INFORMATION

Figure S1. A) Principal component analysis (PCA) between deep Lake Baikal metagenomes based on a Bray-Curtis similarity k-mer profile frequencies of sequencing reads. Red and blue dots represent summer and winter Illumina metagenomes, respectively, while the green dot is the sample retrieved in this study and sequenced with PacBio Sequel II. **B)** Phylum-level composition based on 16S rRNA gene fragments (Illumina and PacBio CCS5 reads) of the different metagenomes. The single metagenome highlighted in green corresponds to the Pacbio sequencing, whilst the rest of datasets belong to previous Illumina sequencing. The phylum Proteobacteria was divided into its class-level classification. Only those groups with abundance values larger than 1% in any of the metagenomes are shown. **C)** Classification of the 1600m PacBio CCS5 16S rRNA reads at a higher taxonomic resolution. Only sequences larger than 1,000 nucleotides were considered. Sequences ascribed to the Patescibacteria phylum are highlighted in green.

Figure S2. Alignment of two LAGs that are complete in a single contig and the respective MAG from the Illumina assembly.

Figure S3. A) Average nucleotide identity based on metagenomic reads (ANlr) of LAGs and the four *Ca. Baikalibacteria* Bins. **B)** ANlr of ten randomly selected sequences of each *Ca. Baikalibacteria* bin.

Figure S4. Maximum likelihood phylogenetic tree of the *Ca. Baikalibacteria* 16S rRNA genes. Sequences outside the deep branch coming from Figure 1 were used as outgroup for the tree. The reads from the four read bins are colored in the figure.

Figure S5. Diversity of 16S rRNA sequences of *Ca. Baikalibacteria* bins. On the left, pair-wise identity matrices of 16S rRNA genes. On the right, linear representation of selected CCS5 reads (indicated with a red circle in the left panel) containing a 16S rRNA gene. A pairwise blastp comparison among reads was performed to detect orthologous genes.

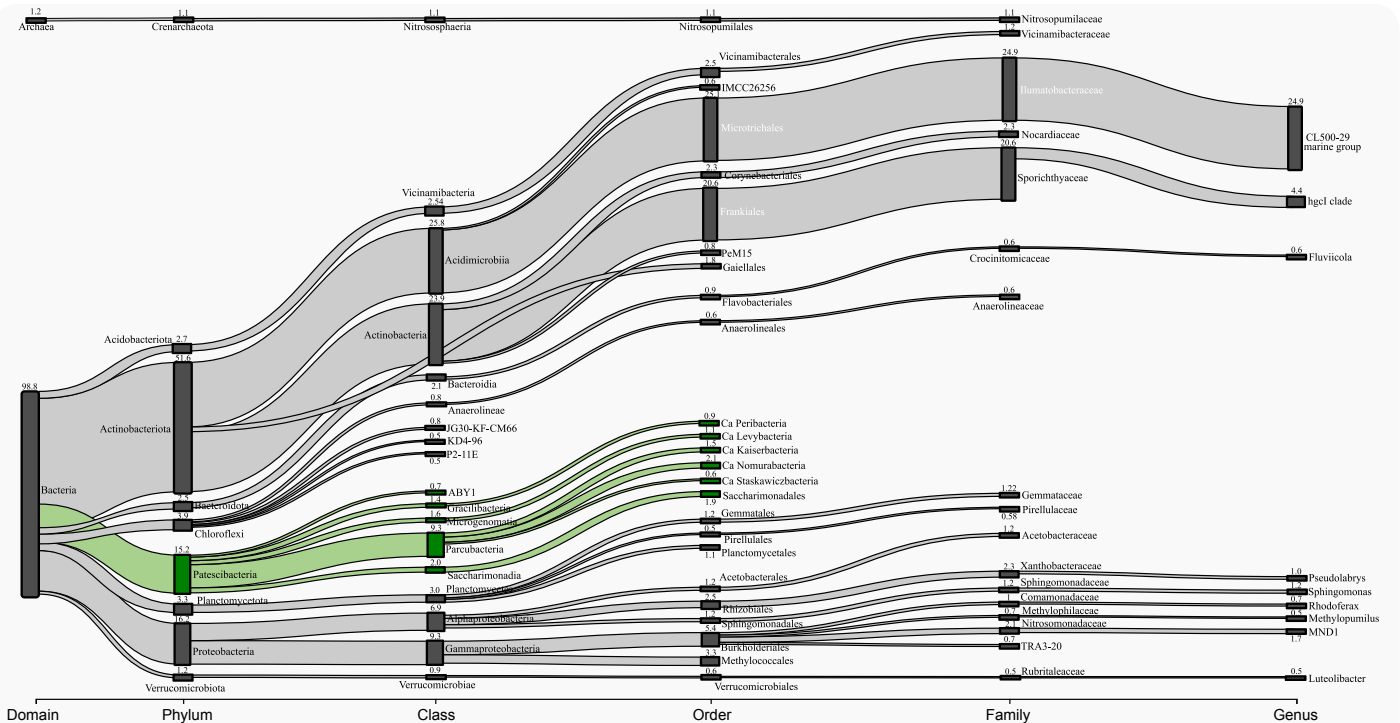
Figure S6. Metagenomic recruitment of the largest fragment of *Ca. Baikalibacteria* RBin09 on Lake Baikal 1600m (upper panel) and Lake Thun 180 m deep (lower panel).

Figure S7. Maximum likelihood phylogenetic tree of the phytoene elongase (LyeJ) protein.

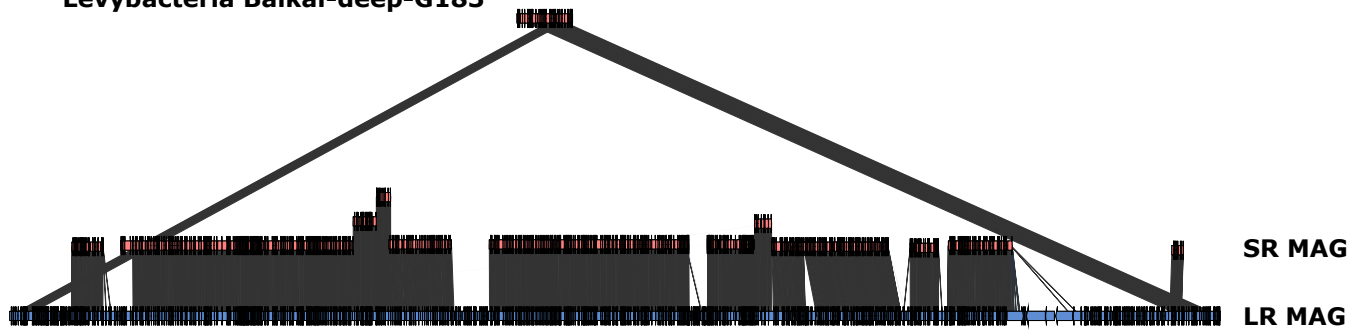
a)



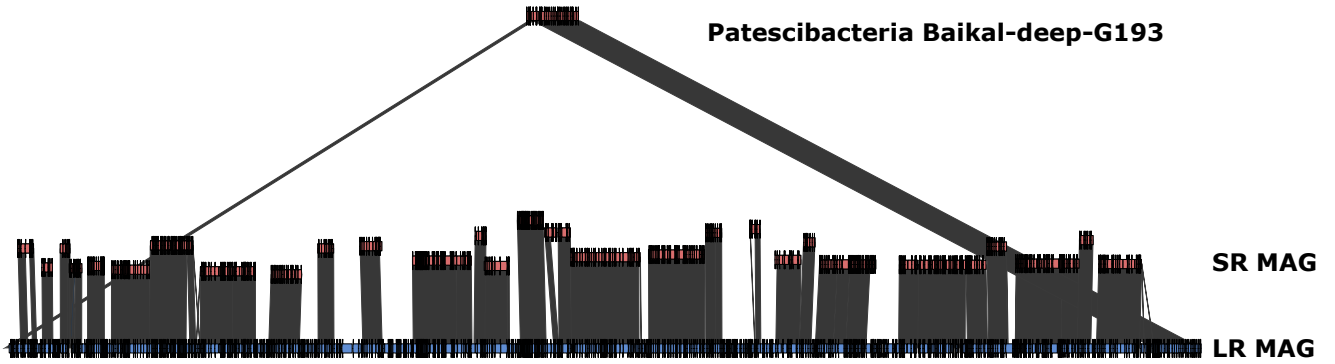
b)

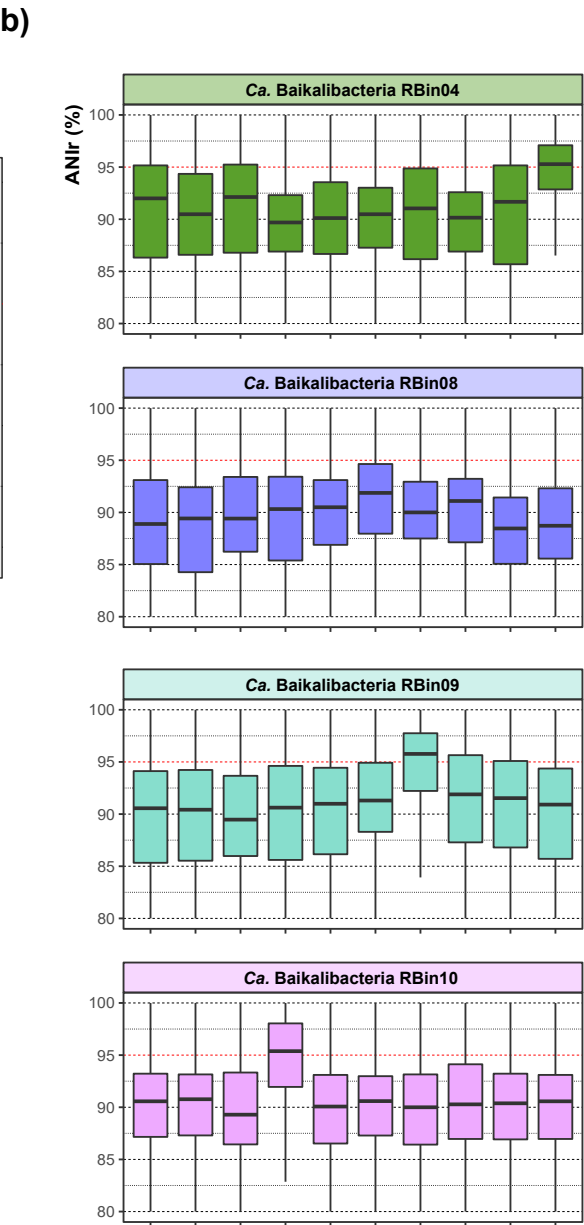
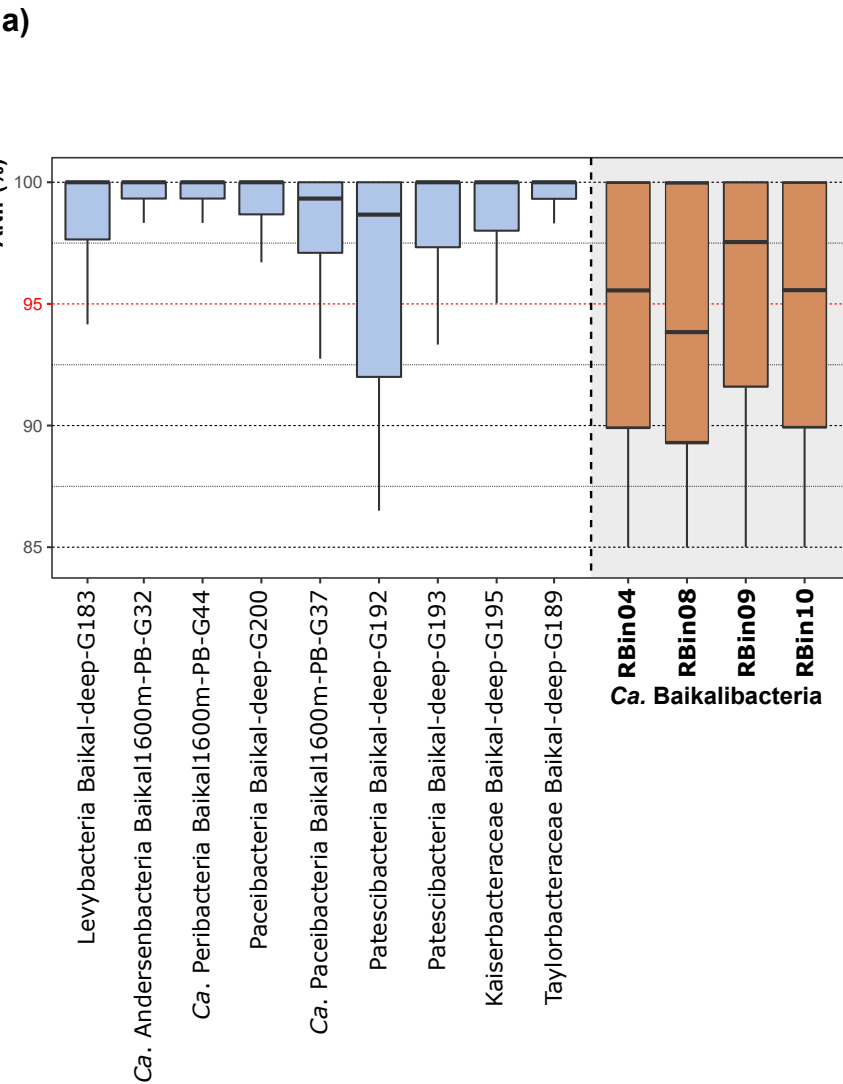


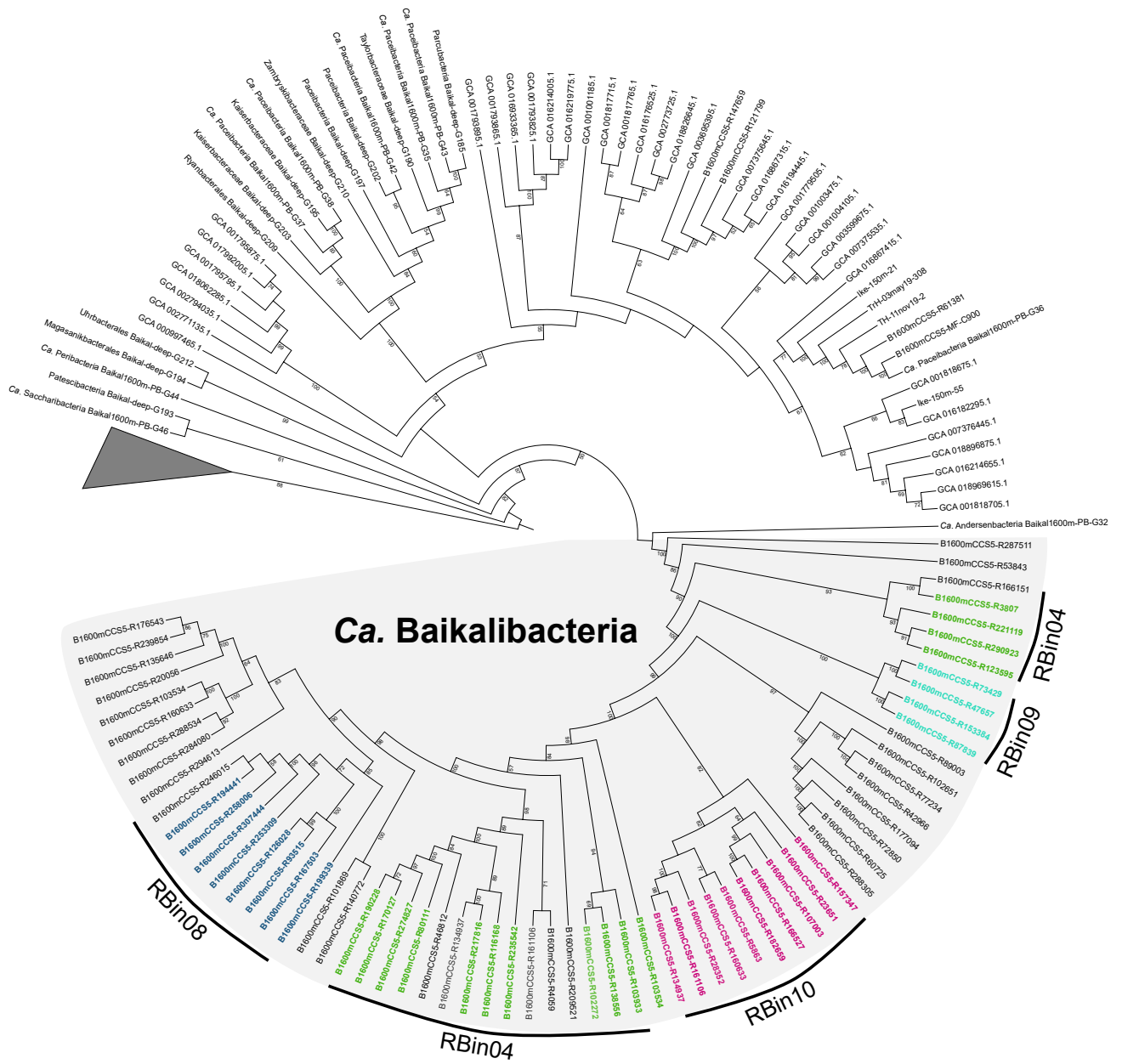
Levybacteria Baikal-deep-G183

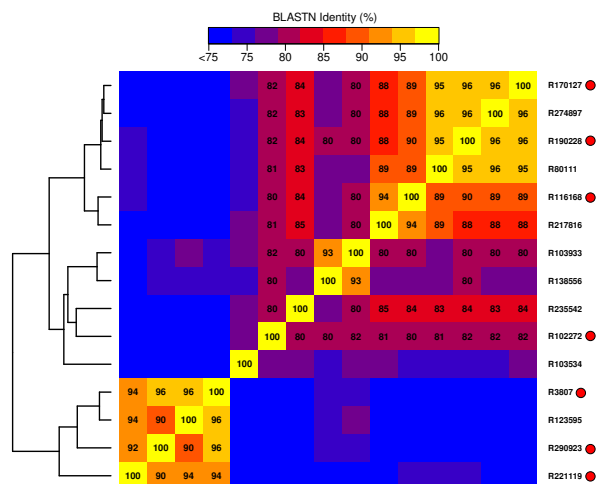


Patescibacteria Baikal-deep-G193

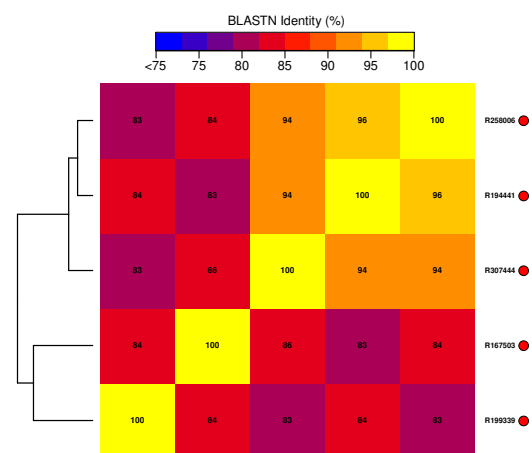
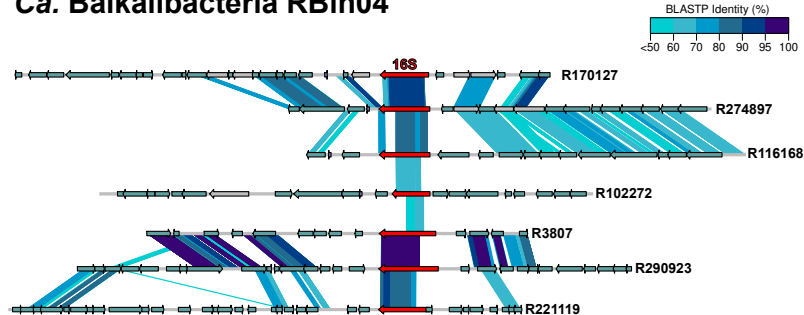




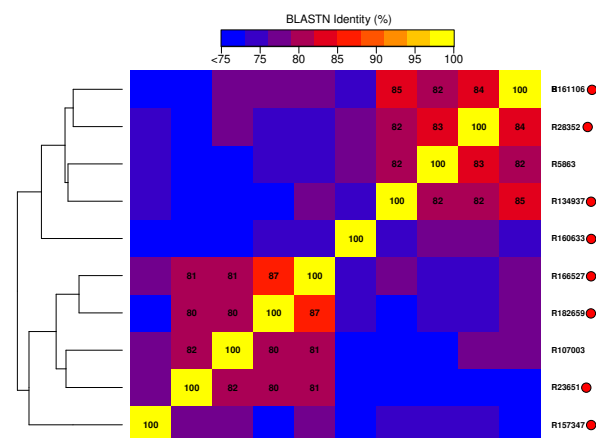
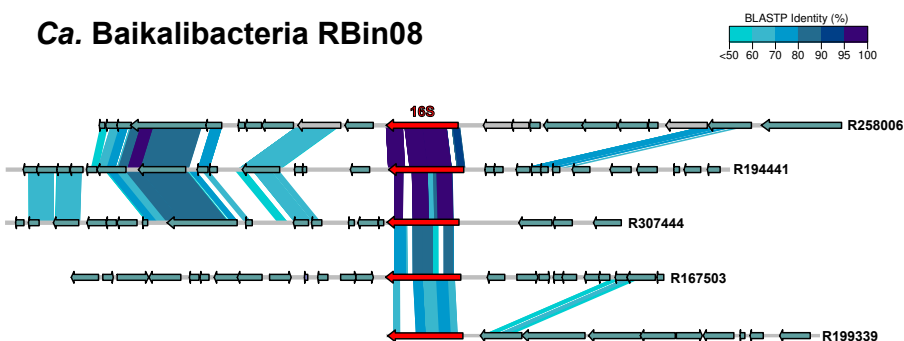




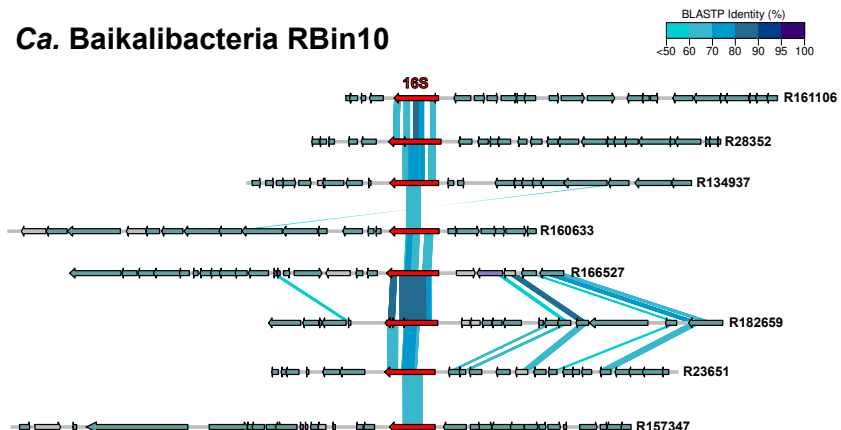
Ca. Baikalibacteria RBin04



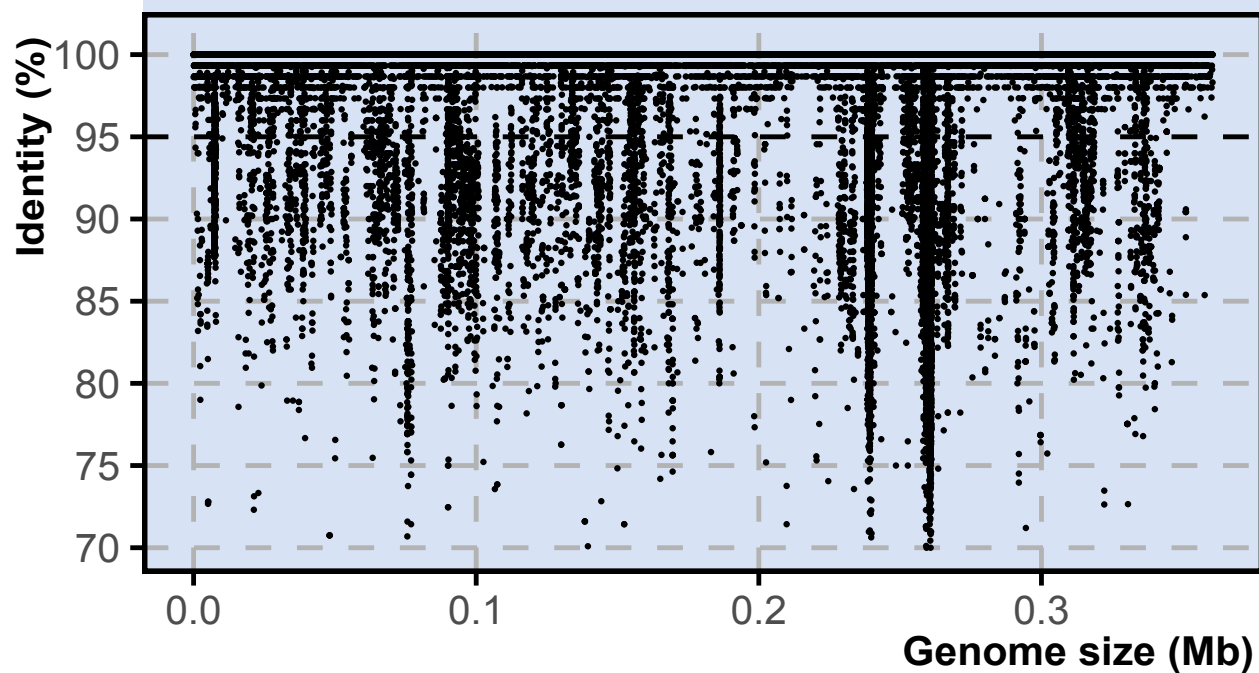
Ca. Baikalibacteria RBin08



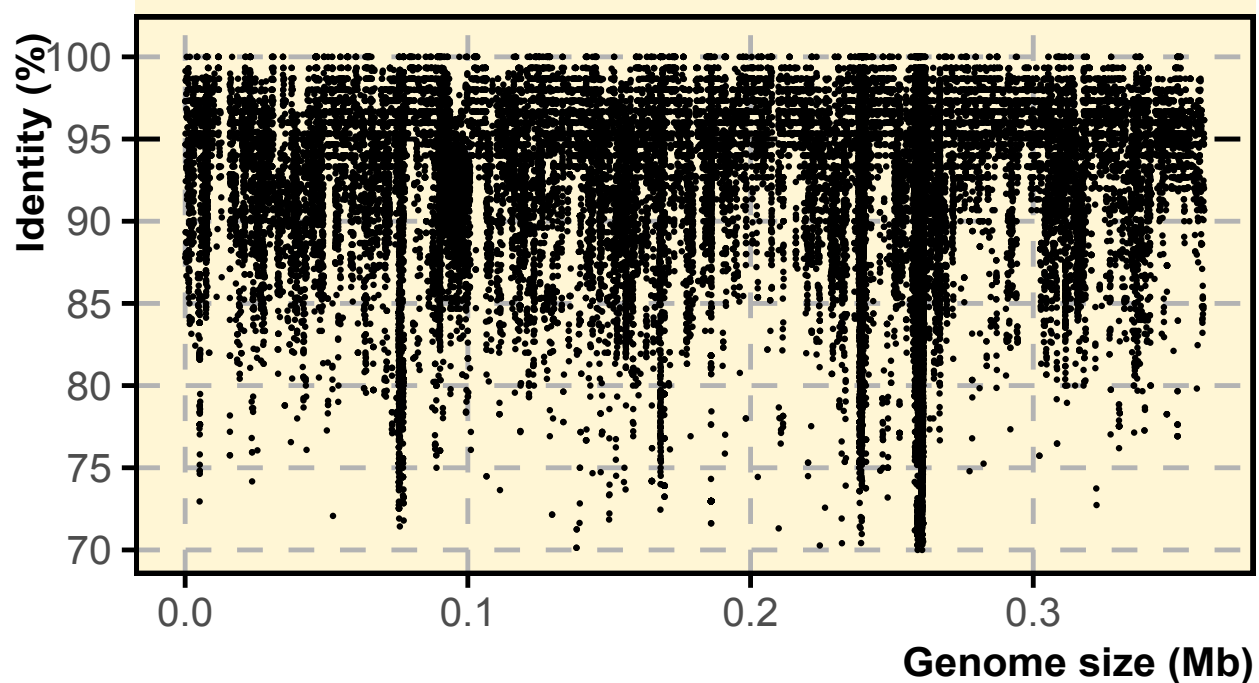
Ca. Baikalibacteria RBin10



Ca. Baikalibacteria RBin09 on Lake Baikal



Ca. Baikalibacteria RBin09 on Lake Thun



lycopene elongase (lyeJ)

