

A new view of the tree of life

Laura A. Hug, Brett J. Baker, Karthik Anantharaman, Christopher T. Brown, Alexander J. Probst, Cindy J. Castelle, Cristina N. Butterfield, Alex W. Hernsdorf, Yuki Amano, Kotaro Ise, Yohey Suzuki, Natasha Dudek, David A. Relman, Kari M. Finstad, Ronald Amundson, Brian C. Thomas, and Jillian F. Banfield*

* Corresponding Author: jbanfield@berkeley.edu

Supplementary Information

Supplementary Figure 1– Full rectangular view of the concatenated ribosomal protein tree presented in Figure 1, with major lineages colored as in the main text Figure 1. Organisms are named with all relevant taxonomic ranks. Bootstrap support values above 50 are displayed on the nodes. This concatenated ribosomal protein tree is available in newick format as Supplementary Dataset 2.

Supplementary Figure 2 – Condensed maximum likelihood tree based on an alignment of SSU rRNA gene sequences. Major lineages are collapsed and labeled as in Figure 1, with well-defined lineages named using italic font and lineages lacking an isolated representative named in regular font and highlighted with red dots. Bootstrap support values are indicated by circles on nodes – black for support of 85% and above, grey for support from 50% to 84%. The full SSU rRNA gene tree is available in newick format as Supplementary Dataset 4.

Supplementary Table 1 - Accession numbers for all genomes included in this study.

Genomes were mined for bacterial-type (P-type) ribosomal proteins L2, L3, L4, L5, L6, L14, L15, L16, L18, L22, L24, S3, S8, S10, S17, and S19 (all single-copy genes), and one representative SSU rRNA gene.

Supplementary Dataset 1 – Concatenated ribosomal protein alignment used for inferring tree in main text Figures 1 and 2, as well as Supplementary Figure 1, in fasta format.

Supplementary Dataset 2 – Maximum likelihood tree based on concatenated ribosomal protein alignment (Supplementary Dataset 1), basis for main text Figures 1 and 2, and Supplementary Figure 1, in newick format.

Supplementary Dataset 3 – SSU rRNA alignment used for inferring tree in Supplementary Figure 2, in fasta format.

Supplementary Dataset 4 – Maximum likelihood tree based on SSU rRNA alignment (Supplementary Dataset 3), basis for Supplementary Figure 2, in newick format.