Publication-ready text:

METHODS

Pairwise sequence similarities were calculated using the method recommended by Meier-Kolthoff et al. (2013) for the 16S rRNA gene available via the GGDC web server (Meier-Kolthoff et al. 2021) available at http://ggdc.dsmz.de/.

Phylogenies were inferred by the GGDC web server (Meier-Kolthoff et al. 2021) available at http://ggdc.dsmz.de/ using the DSMZ phylogenomics pipeline (Meier-Kolthoff et al. 2014) adapted to single genes. A multiple sequence alignment was created with MUSCLE (Edgar 2004). Maximum likelihood (ML) and maximum parsimony (MP) trees were inferred from the alignment with RAxML (Stamatakis 2014) and TNT (Goloboff et al. 2008), respectively. For ML, rapid bootstrapping in conjunction with the autoMRE bootstopping criterion (Pattengale et al. 2010) and subsequent search for the best tree was used; for MP, 1000 bootstrapping replicates were used in conjunction with tree-bisection-and-reconnection branch swapping and ten random sequence addition replicates. The sequences were checked for a compositional bias using the Χ² test as implemented in PAUP\* (Swofford 2002).

RESULTS

The input nucleotide matrix comprised 52 operational taxonomic units and 1558 characters, 529 of which were variable and 420 of which were parsimony-informative. The base-frequency check indicated no compositional bias (p = 1.00, α = 0.05). ML analysis under the GTR+CAT model yielded a highest log likelihood of -10386.56, whereas the estimated alpha parameter was 0.21. The ML bootstrapping converged after 600 replicates; the average support was 51.92%. MP analysis yielded a best score of 1726 (consistency index 0.46, retention index 0.72) and 20 best trees. The MP bootstrapping average support was 64.76%.

Figure 1 [adapt the number to your manuscript!]. ML tree inferred under the GTR+CAT model and rooted by midpoint-rooting [change this if the tree was re-rooted!]. The branches are scaled in terms of the expected number of substitutions per site. The numbers above the branches are support values when larger than 60% from ML (left) and MP (right) bootstrapping.

REFERENCES

Edgar RC (2004) MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res 32: 1792-1797 (doi:10.1093/nar/gkh340).

Goloboff PA, Farris JS, Nixon KC (2008) TNT, a free program for phylogenetic analysis. Cladistics 24: 774-786 (doi:10.1111/j.1096-0031.2008.00217.x).

Hess PN, De Moraes Russo CA (2007). An empirical test of the midpoint rooting method. Biol J Linn Soc 92: 669-674 (doi:10.1111/j.1095-8312.2007.00864.x).

Meier-Kolthoff JP, Carbasse JS, Peinado-Olarte RL, Göker M (2021) TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. Nucleic Acids Res (doi:10.1186/1471-2105-14-60).

Meier-Kolthoff JP, Göker M, Spröer C, Klenk H-P (2013) When should a DDH experiment be mandatory in microbial taxonomy? Arch Microbiol 195: 413-418 (doi:10.1007/s00203-013-0888-4).

Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, Rohde C, Rohde M, Fartmann B, Goodwin LA, Chertkov O, Reddy T, Pati A, Ivanova N, Markowitz V, Kyrpides NC, Woyke T, Göker M, Klenk H-P (2014) Complete genome sequence of DSM 30083T, the type strain (U5/41T) of Escherichia coli, and a proposal for delineating subspecies in microbial taxonomy. Stand Genomic Sci 10: 2, 2014 (doi:10.1186/1944-3277-9-2).

Pattengale ND, Alipour M, Bininda-Emonds ORP, Moret BME, Stamatakis A (2010) How many bootstrap replicates are necessary? J Comput Biol 17: 337–354.

Stamatakis A (2014) RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30: 1312-1313 (doi:10.1093/bioinformatics/btu033).

Swofford DL (2002) PAUP\*: Phylogenetic Analysis Using Parsimony (\*and Other Methods), Version 4.0 b10. Sinauer Associates, Sunderland.