



PRINT DATE: 2023-07-26 15:24:47 +0200

JOB ID: d9fe4271-8a4e-4e2c-839b-fabf74d921b7

RESULT PAGE: https://tygs.dsmz.de/user_results/show?guid=d9fe4271-8a4e-4e2c-839b-fabf74d921b7

Table 1: Phylogenies

Publication-ready versions of both the genome-scale GBDP tree and the 16S rRNA gene sequence tree can be customized and exported either in SVG (vector graphic) or PNG format from within the phylogeny viewers in your TYGS result page. For publications the **SVG format is recommended** because it is lossless, always keeps its high resolution and can also be easily converted to other popular formats such as PDF or EPS. Please follow the link provided above!

Table 2: Identification

Note: the identification of your genomes is not possible in the currently chosen mode because the analysis was restricted to user genomes only.

Table 3: Pairwise comparisons of user genomes vs. type-strain genomes

The following table contains the pairwise dDDH values between your user genomes and the selected type-strain genomes. The dDDH values are provided along with their confidence intervals (C.I.) for the three different GBDP formulas:

- formula d_0 (a.k.a. GGDC formula 1): length of all HSPs divided by total genome length
- formula d_4 (a.k.a. GGDC formula 2): sum of all identities found in HSPs divided by overall HSP length
- formula d_6 (a.k.a. GGDC formula 3): sum of all identities found in HSPs divided by total genome length

Note: Formula d_4 is independent of genome length and is thus robust against the use of incomplete draft genomes. For other reasons for preferring formula d_4 , see the FAQ.

Query	Subject	d_0	C.I. d_0	d_4	C.I. d_4	d_6	C.I. d_6	Diff. G+C Percent
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_001267215.1_ASM126721v1_genomic.fna'	94.3	[91.8 - 96.1]	96.7	[95.4 - 97.6]	96.5	[94.9 - 97.6]	0.05
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_001267255.1_ASM126725v1_genomic.fna'	91.9	[88.9 - 94.1]	94.5	[92.8 - 95.8]	94.5	[92.5 - 96.0]	0.03
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_001267255.1_ASM126725v1_genomic.fna'	95.0	[92.7 - 96.6]	94.5	[92.8 - 95.8]	96.7	[95.2 - 97.7]	0.08
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_001267215.1_ASM126721v1_genomic.fna'	89.7	[86.5 - 92.3]	93.7	[91.8 - 95.1]	92.9	[90.5 - 94.7]	0.14
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_001267195.1_ASM126719v1_genomic.fna'	85.2	[81.4 - 88.2]	93.1	[91.2 - 94.6]	89.3	[86.4 - 91.6]	0.1
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_001267255.1_ASM126725v1_genomic.fna'	86.2	[82.5 - 89.2]	92.7	[90.7 - 94.2]	90.0	[87.3 - 92.2]	0.07
'GCF_001267315.1_ASM126731v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	83.1	[79.3 - 86.4]	91.1	[88.9 - 92.9]	87.4	[84.3 - 89.9]	0.1
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_001267275.1_ASM126727v1_genomic.fna'	90.6	[87.4 - 93.0]	89.6	[87.2 - 91.5]	92.9	[90.6 - 94.7]	0.12
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_001267275.1_ASM126727v1_genomic.fna'	87.5	[84.0 - 90.3]	89.6	[87.3 - 91.6]	90.6	[87.9 - 92.7]	0.08

Query	Subject	d_0	C.I. d_0	d_4	C.I. d_4	d_6	C.I. d_6	Diff. G+C Percent
'GCF_001267255.1_ASM126725v1_genomic.fna'	'GCF_001267275.1_ASM126727v1_genomic.fna'	89.4	[86.1 - 92.0]	89.5	[87.1 - 91.5]	92.0	[89.6 - 94.0]	0.04
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_001267275.1_ASM126727v1_genomic.fna'	83.1	[79.3 - 86.3]	89.2	[86.8 - 91.2]	87.0	[84.0 - 89.6]	0.02
'GCF_001267275.1_ASM126727v1_genomic.fna'	'GCF_001267295.1_ASM126729v1_genomic.fna'	85.5	[81.8 - 88.5]	89.1	[86.7 - 91.1]	88.9	[86.1 - 91.3]	0.06
'GCF_001267255.1_ASM126725v1_genomic.fna'	'GCF_001267295.1_ASM126729v1_genomic.fna'	89.1	[85.7 - 91.7]	88.6	[86.1 - 90.7]	91.7	[89.1 - 93.6]	0.02
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_001267295.1_ASM126729v1_genomic.fna'	90.8	[87.6 - 93.2]	88.6	[86.1 - 90.6]	92.9	[90.6 - 94.7]	0.06
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_001267295.1_ASM126729v1_genomic.fna'	87.0	[83.4 - 89.9]	88.4	[86.0 - 90.5]	90.0	[87.2 - 92.2]	0.01
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_001267295.1_ASM126729v1_genomic.fna'	91.0	[87.9 - 93.3]	88.3	[85.9 - 90.4]	93.0	[90.7 - 94.8]	0.08
'GCF_001267255.1_ASM126725v1_genomic.fna'	'GCF_001267315.1_ASM126731v1_genomic.fna'	84.2	[80.4 - 87.4]	86.8	[84.2 - 89.0]	87.5	[84.5 - 90.0]	0.22
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	79.7	[75.8 - 83.2]	86.7	[84.1 - 88.9]	83.8	[80.6 - 86.6]	0.36
'GCF_001267255.1_ASM126725v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	82.9	[79.1 - 86.2]	86.7	[84.1 - 88.9]	86.5	[83.4 - 89.1]	0.32
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	84.5	[80.7 - 87.6]	86.6	[83.9 - 88.8]	87.7	[84.7 - 90.2]	0.4
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_001267315.1_ASM126731v1_genomic.fna'	87.8	[84.3 - 90.6]	86.6	[84.0 - 88.9]	90.4	[87.7 - 92.6]	0.3
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_001267315.1_ASM126731v1_genomic.fna'	84.1	[80.3 - 87.3]	86.5	[83.8 - 88.7]	87.4	[84.4 - 89.9]	0.16
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_001267315.1_ASM126731v1_genomic.fna'	85.0	[81.3 - 88.1]	86.4	[83.7 - 88.6]	88.1	[85.2 - 90.5]	0.26
'GCF_001267275.1_ASM126727v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	82.0	[78.1 - 85.3]	85.9	[83.2 - 88.2]	85.6	[82.4 - 88.2]	0.28
'GCF_001267275.1_ASM126727v1_genomic.fna'	'GCF_001267315.1_ASM126731v1_genomic.fna'	84.0	[80.2 - 87.2]	85.9	[83.2 - 88.2]	87.3	[84.2 - 89.8]	0.18
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	77.8	[73.8 - 81.3]	85.7	[83.0 - 88.0]	82.0	[78.7 - 85.0]	0.26
'GCF_001267295.1_ASM126729v1_genomic.fna'	'GCF_001267315.1_ASM126731v1_genomic.fna'	84.8	[81.0 - 87.9]	85.4	[82.7 - 87.7]	87.8	[84.8 - 90.2]	0.24
'GCF_001267295.1_ASM126729v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	83.3	[79.4 - 86.5]	85.2	[82.5 - 87.5]	86.5	[83.4 - 89.1]	0.34
'GCF_001267315.1_ASM126731v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	81.2	[77.3 - 84.6]	80.4	[77.5 - 83.1]	84.0	[80.8 - 86.8]	0.4
'GCF_030328955.1_ASM3032895v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	76.7	[72.7 - 80.2]	79.7	[76.8 - 82.4]	80.0	[76.6 - 83.1]	0.5
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	78.4	[74.4 - 81.9]	79.0	[76.0 - 81.7]	81.4	[78.0 - 84.3]	0.1
'GCF_001267255.1_ASM126725v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	76.5	[72.5 - 80.1]	78.7	[75.7 - 81.4]	79.7	[76.3 - 82.7]	0.18
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	71.9	[67.9 - 75.5]	78.6	[75.6 - 81.3]	75.6	[72.1 - 78.8]	0.24

Query	Subject	d_0	C.I. d_0	d_4	C.I. d_4	d_6	C.I. d_6	Diff. G+C Percent
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	74.4	[70.5 - 78.1]	78.1	[75.1 - 80.8]	77.8	[74.3 - 80.9]	0.15
'GCF_001267275.1_ASM126727v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	78.2	[74.2 - 81.7]	77.8	[74.8 - 80.5]	81.0	[77.6 - 84.0]	0.22
'GCF_001267295.1_ASM126729v1_genomic.fna'	'GCF_030328955.1_ASM3032895v1_genomic.fna'	75.6	[71.6 - 79.1]	77.4	[74.4 - 80.2]	78.7	[75.2 - 81.7]	0.16
'GCF_030328955.1_ASM3032895v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	77.7	[73.7 - 81.2]	65.7	[62.8 - 68.5]	78.1	[74.7 - 81.2]	0.55
'GCF_001267315.1_ASM126731v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	77.3	[73.3 - 80.8]	65.4	[62.4 - 68.2]	77.7	[74.2 - 80.8]	0.96
'GCF_030328995.1_ASM3032899v1_genomic.fna'	'GCF_030329025.1_ASM3032902v1_genomic.fna'	74.1	[70.2 - 77.8]	64.7	[61.8 - 67.6]	74.8	[71.4 - 78.0]	1.05
'GCF_001267175.1_ASM126717v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	71.5	[67.5 - 75.1]	64.3	[61.4 - 67.1]	72.4	[68.9 - 75.6]	0.8
'GCF_001267275.1_ASM126727v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	76.3	[72.4 - 79.9]	63.9	[61.0 - 66.8]	76.5	[73.1 - 79.7]	0.77
'GCF_001267295.1_ASM126729v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	75.6	[71.6 - 79.2]	63.8	[60.9 - 66.6]	75.9	[72.4 - 79.1]	0.71
'GCF_001267215.1_ASM126721v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	77.7	[73.7 - 81.3]	63.8	[60.9 - 66.6]	77.7	[74.2 - 80.8]	0.65
'GCF_001267255.1_ASM126725v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	78.2	[74.2 - 81.7]	63.7	[60.8 - 66.6]	78.0	[74.6 - 81.1]	0.73
'GCF_001267195.1_ASM126719v1_genomic.fna'	'GCF_030328995.1_ASM3032899v1_genomic.fna'	76.1	[72.2 - 79.7]	63.0	[60.1 - 65.8]	76.2	[72.7 - 79.3]	0.7

Table 4: Strains in your dataset

Joint dataset of automatically determined closest type strains (if this mode was chosen), manually selected type strains (if selected accordingly) and the provided user strains, if provided (marked in **yellow**).

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
GCF_001267175 .1_ASM126717v 1_genomic.fna				2494 031	47.4	2366					
GCF_001267195 .1_ASM126719v 1_genomic.fna				2272 143	47.5	2141					
GCF_001267215 .1_ASM126721v 1_genomic.fna				2312 033	47.5	2158					
GCF_001267255 .1_ASM126725v 1_genomic.fna				2294 397	47.4	2139					
GCF_001267275 .1_ASM126727v 1_genomic.fna				2391 994	47.4	2218					
GCF_001267295 .1_ASM126729v 1_genomic.fna				2377 259	47.4	2235					
GCF_001267315 .1_ASM126731v 1_genomic.fna				2397 029	47.2	2269					
GCF_030328955 .1_ASM3032895 v1_genomic.fna				2572 910	47.6	2431					
GCF_030328995 .1_ASM3032899 v1_genomic.fna				2475 608	48.2	2344					
GCF_030329025 .1_ASM3032902 v1_genomic.fna				2440 383	47.1	2368					

Methods, Results and References

The genome sequence data were uploaded to the Type (Strain) Genome Server (TYGS), a free bioinformatics platform available under <https://tygs.dsmz.de>, for a whole genome-based taxonomic analysis [1]. The analysis also made use of recently introduced methodological updates and features [2]. Information on nomenclature, synonymy and associated taxonomic literature was provided by TYGS's sister database, the List of Prokaryotic names with Standing in Nomenclature (LPSN, available at <https://lpsn.dsmz.de>) [2]. The results were provided by the TYGS on 2023-07-25. The TYGS analysis was subdivided into the following steps:

Determination of closely related type strains

Determination of closest type strain genomes was done in two complementary ways: First, all user genomes were compared against all type strain genomes available in the TYGS database via the MASH algorithm, a fast approximation of intergenomic relatedness [3], and, the ten type strains with the smallest MASH distances chosen per user genome. Second, an additional set of ten closely related type strains was determined via the 16S rDNA gene sequences. These were extracted from the user genomes using RNAmmer [4] and each sequence was subsequently BLASTed [5] against the 16S rDNA gene sequence of each of the currently 19258 type strains available in the TYGS database. This was used as a proxy to find the best 50 matching type strains (according to the bitscore) for each user genome and to subsequently calculate precise distances using the Genome BLAST Distance Phylogeny approach (GBDP) under the algorithm 'coverage' and distance formula d_5 [6]. These distances were finally used to determine the 10 closest type strain genomes for each of the user genomes.

Pairwise comparison of genome sequences

For the phylogenomic inference, all pairwise comparisons among the set of genomes were conducted using GBDP and accurate intergenomic distances inferred under the algorithm 'trimming' and distance formula d_5 [6]. 100 distance replicates were calculated each. A second GBDP phylogenomic analysis was inferred using the amino acid sequences of the entire proteome, promising a better resolved phylogeny in datasets of only remotely related strains. Digital DDH values and confidence intervals were calculated using the recommended settings of the GGDC 3.0 [2,6].

Phylogenetic inference

The resulting intergenomic distances were used to infer a balanced minimum evolution tree with branch support via FASTME 2.1.6.1 including SPR postprocessing [7]. Branch support was inferred from 100 pseudo-bootstrap replicates each. The trees were rooted at the midpoint [8] and visualized with PhyD3 [9].

Type-based species and subspecies clustering

The type-based species clustering using a 70% dDDH radius around each of the 0 type strains was done as previously described [1]. The resulting groups are shown in Table 1 and 4. Subspecies clustering was done using a 79% dDDH threshold as previously introduced [10].

Results

Type-based species and subspecies clustering

The resulting species and subspecies clusters are listed in Table 4, whereas the taxonomic identification of the query strains is found in Table 1. Briefly, the clustering yielded species clusters and the provided query strains were assigned to of these. Moreover, user strains were located in of subspecies clusters.

Figure caption SSU tree

Figure 1. Tree inferred with FastME 2.1.6.1 [7] from GBDP distances calculated from 16S rDNA gene sequences. The branch lengths are scaled in terms of GBDP distance formula d_5 . The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 45.4 %. The tree was rooted at the midpoint [8].

Figure caption genome tree

Figure 2. Tree inferred with FastME 2.1.6.1 [7] from GBDP distances calculated from genome sequences. The branch lengths are scaled in terms of GBDP distance formula d_5 . The numbers above branches are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 19.3 %. The tree was rooted at the midpoint [8].

Figure caption proteome tree

Figure 3. Tree inferred with FastME 2.1.6.1 [7] from whole-proteome-based GBDP distances. The branch lengths are scaled via GBDP distance formula d_5 . Branch values are GBDP pseudo-bootstrap support values > 60 % from 100 replications, with an average branch support of 57.3 %. The tree was midpoint-rooted [8].

References

- [1] Meier-Kolthoff JP, Göker M. TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy. *Nat. Commun.* 2019;10: 2182. DOI: 10.1038/s41467-019-10210-3
- [2] Meier-Kolthoff JP, Sardà Carbasse J, Peinado-Olarte RL, Göker M. TYGS and LPSN: a database tandem for fast and reliable genome-based classification and nomenclature of prokaryotes. *Nucleic Acid Res.* 2022;50: D801–D807. DOI: 10.1093/nar/gkab902
- [3] Ondov BD, Treangen TJ, Melsted P, et al. Mash: Fast genome and metagenome distance estimation using MinHash. *Genome Biol* 2016;17: 1–14. DOI: 10.1186/s13059-016-0997-x
- [4] Lagesen K, Hallin P. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res. Oxford Univ Press*; 2007;35: 3100–3108. DOI: 10.1093/nar/gkm160
- [5] Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics.* 2009;10: 421. DOI: 10.1186/1471-2105-10-421
- [6] Meier-Kolthoff JP, Auch AF, Klenk H-P, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics.* 2013;14: 60. DOI: 10.1186/1471-2105-14-60
- [7] Lefort V, Desper R, Gascuel O. FastME 2.0: A comprehensive, accurate, and fast distance-based phylogeny inference program. *Mol Biol Evol.* 2015;32: 2798–2800. DOI: 10.1093/molbev/msv150
- [8] Farris JS. Estimating phylogenetic trees from distance matrices. *Am Nat.* 1972;106: 645–667.
- [9] Kreft L, Botzki A, Coppens F, Vandepoele K, Van Bel M. PhyD3: A phylogenetic tree viewer with extended phyloXML support for functional genomics data visualization. *Bioinformatics.* 2017;33: 2946–2947. DOI: 10.1093/bioinformatics/btx324
- [10] Meier-Kolthoff JP, Hahnke RL, Petersen J, Scheuner C, Michael V, Fiebig A, et al. Complete genome sequence of DSM 30083^T, the type strain (U5/41^T) of *Escherichia coli*, and a proposal for delineating subspecies in microbial taxonomy. *Stand Genomic Sci.* 2014;9: 2. DOI: 10.1186/1944-3277-9-2