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

The below list contains the result of the TYGS species identification routine.

Explanation of remarks that might occur in the below table:

remark [R1]: The TYGS type strain database is automatically updated on an almost daily basis. However, if a particular type strain genome is not available in the TYGS database, this can have several reasons which are detailed in the FAQ. You can request an extended 16S rRNA gene analysis via the 16S tree viewer found in your result page to detect **not yet genome-sequenced** type strains relevant for your study.

remark [R2]: > 70% dDDH value (formula d_4) and (almost) minimal dDDH values for gene-content formulae d_0 and d_6 indicate a potentially unreliable identification result and should thus be checked via the 16S rRNA gene sequence similarity. Such strong deviations can, in principle, be caused by sequence contamination.

remark [R3]: G+C content difference of > 1 % indicates a potentially unreliable identification result because within species G+C content varies no more than 1 %, if computed from genome sequences (PMID: 24505073).

Strain	Conclusion	Identification result	Remark
Azotobacter	belongs to known species	*Azotobacter beijerinckii*	

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
'Azotobacter'	*Azotobacter beijerinckii* DSM 378	79.7	[75.7 - 83.1]	72.0	[68.9 - 74.8]	81.2	[77.8 - 84.1]	0.3
'Azotobacter'	*Azotobacter chroococcum* ATCC 9043	53.1	[49.6 - 56.6]	41.5	[39.0 - 44.0]	50.7	[47.6 - 53.7]	0.24
'Azotobacter'	*Azotobacter chroococcum* DSM 2286	53.0	[49.6 - 56.5]	41.4	[38.9 - 44.0]	50.6	[47.5 - 53.7]	0.25
'Azotobacter'	*Azotobacter chroococcum* subsp. *isscasi* P205	54.3	[50.8 - 57.8]	41.1	[38.6 - 43.7]	51.6	[48.5 - 54.6]	0.86
'Azotobacter'	*Pseudomonas linyingensis* LMG 25967	23.5	[20.2 - 27.2]	24.9	[22.6 - 27.3]	22.6	[19.8 - 25.7]	0.49
'Azotobacter'	*Pseudomonas sagittaria* JCM 18195	23.4	[20.1 - 27.0]	24.9	[22.6 - 27.4]	22.5	[19.7 - 25.6]	0.76
'Azotobacter'	*Pseudomonas oryzae* KCTC 32247	24.0	[20.7 - 27.6]	24.9	[22.6 - 27.4]	23.0	[20.2 - 26.1]	1.45
'Azotobacter'	*Pseudomonas guangdongensis* CCTCCAB2012022	23.0	[19.8 - 26.7]	24.4	[22.1 - 26.9]	22.2	[19.4 - 25.3]	2.37
'Azotobacter'	*Pseudomonas lalucatii* R1b-54T	22.0	[18.8 - 25.7]	24.3	[22.0 - 26.8]	21.4	[18.6 - 24.4]	1.1
'Azotobacter'	*Pseudomonas schmalbachii* BCRC 81294	20.0	[16.8 - 23.6]	22.8	[20.5 - 25.3]	19.5	[16.8 - 22.6]	0.93
'Azotobacter'	*Pseudomonas panipatensis* CCM 7469	19.8	[16.6 - 23.4]	22.6	[20.3 - 25.0]	19.3	[16.6 - 22.3]	0.34
'Azotobacter'	*Pseudomonas indoloxydans* JCM 14246	18.2	[15.1 - 21.7]	22.5	[20.3 - 25.0]	18.0	[15.3 - 20.9]	3.65
'Azotobacter'	*Pseudomonas pseudoalcaligenes* NBRC 14167	18.6	[15.5 - 22.2]	22.4	[20.2 - 24.9]	18.3	[15.6 - 21.3]	3.65
'Azotobacter'	*Pseudomonas mangrovi* TC11	19.6	[16.4 - 23.2]	22.4	[20.1 - 24.8]	19.1	[16.4 - 22.1]	1.56
'Azotobacter'	*Pseudomonas toyotomiensis* JCM 15604	18.4	[15.3 - 22.0]	22.1	[19.9 - 24.6]	18.1	[15.4 - 21.1]	3.29

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
Pseudomonas sagittaria JCM 18195	Lin et al. 2013	BCRC 80399; DSM 27945; CC-OPY-1	*Pseudomonas sagittaria*	4607 240	66.7	4197	Gp0127211	PRJEB17413	SAMN05216229	GCA_900115715	
Pseudomonas toyotomiensis JCM 15604	Hirota et al. 2011	NCIMB 14511; DSM 26169; HT-3	*Pseudomonas toyotomiensis*	5489 804	62.6	5160	Gp0127184	PRJEB17410	SAMN05216177	GCA_900115695	
Azotobacter chroococcum subsp. *isscasi* P205	Jin et al. 2020	KCTC 72233; CGMCC 1.16846; CCTCC AB 2019080	*Azotobacter chroococcum* subsp. *isscasi*	4638 648	66.8	4117		PRJNA523901	SAMN10992292	GCA_004327895	
Pseudomonas mangrovi TC11	Ye et al. 2019	MCCC 1K03499; KCTC 62159	*Pseudomonas mangrovi*	4054 179	64.3	3692	Gp0388945	PRJNA449190	SAMN08891238	GCA_003052585	
Pseudomonas lalucatii R1b-54T	Busquets et al. 2021	CCUG 74754T; CECT 30179T	*Pseudomonas lalucatii*	4583 128	67.0	4133		PRJNA678722	SAMN16810909	GCA_018398425	
Pseudomonas indoloxydans JCM 14246	Manickam et al. 2008	IPL-1; MTCC 8062	*Pseudomonas indoloxydans*	5198 577	62.2	4666	Gp0323509	PRJNA449191	SAMN08891239	GCA_003052605	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Azotobacter chroococcum DSM 2286	Beijerinck 1901 emend. Jin et al. 2020	LMG 8756; NRRL B-14346; NRRL B-14637; ATCC 9043; DSM 2286; JCM 20725; JCM 21503; NBRC 102613; VKM B-1616; IAM 12666; NCAIM B.01391	*Azotobacter chroococcum*; *Azotobacter chroococcum* subsp. *chroococcum*	4854148	66.2	4515	Gp0325748	PRJNA520333	SAMN10866327	GCA_004339665	2802428852
Pseudomonas pseudoalcaligenes NBRC 14167	Stanier 1966	CFBP 2435; LMG 1225; NCCB 76045; ATCC 17440; CCUG 51525; DSM 50188; JCM 5968; IFO 14167; IFO (now NBRC) 14167; NCTC 10860; CIP 66.14	*Pseudomonas pseudoalcaligenes*; *Pseudomonas pseudoalcaligenes* subsp. *pseudoalcaligenes*	4702414	62.2	4507	Gp0023768	PRJDB224	SAMD00046936	GCA_002091775	
Azotobacter beijerinckii DSM 378	Lipman 1904	LMG 1265; NRRL B-14367; NRRL B-14640; ICMP 8673; ATCC 19360; JCM 20742; VKM B-1615; CIP 106282; NCAIM B.01800	*Azotobacter beijerinckii*	4936268	65.6	4597	Gp0108301	PRJNA297913	SAMN04244573	GCA_900110885	2597490340

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Pseudomonas panipatensis CCM 7469	Gupta et al. 2008	DSM 21819; MTCC 8990; strain Esp-1	*Pseudomonas panipatensis*	5775 045	65.6	5273	Gp0127172	PRJEB15946	SAMN05216272	GCA_900099785	
Pseudomonas guangdongensis CCTCCAB201202 2	Yang et al. 2013	DSM 100318; KACC 16606; SgZ-6	*Pseudomonas guangdongensis*	3191 054	68.3	2892	Gp0127198	PRJNA224116	SAMN05216580	GCF_900105885	
Azotobacter chroococcum ATCC 9043	Beijerinck 1901 emend. Jin et al. 2020	LMG 8756; NRRL B-14346; NRRL B-14637; ATCC 9043; DSM 2286; JCM 20725; JCM 21503; NBRC 102613; VKM B-1616; IAM 12666; NCAIM B.01391	*Azotobacter chroococcum*; *Azotobacter chroococcum* subsp. *chroococcum*	4870 808	66.1	4488		PRJNA224116	SAMN10992294	GCF_004327905	
Pseudomonas linyingensis LMG 25967	He et al. 2015	CGMCC 1.10701; LYBRD3-7	*Pseudomonas linyingensis*	4718 048	66.4	4335	Gp0127205	PRJEB16805	SAMN05216201	GCA_900109175	
Pseudomonas schmalbachii BCRC 81294	Shelomi et al. 2021	JCM 34414; CIP 111980; Milli4	*Pseudomonas schmalbachii*	5464 051	65.0	4730		PRJNA687116	SAMN17134837	GCA_017589465	
Pseudomonas oryzae KCTC 32247	Yu et al. 2013	CGMCC 1.12417; WM-3	*Pseudomonas oryzae*	4642 093	67.4	4110	Gp0127209	PRJEB16377	SAMN05216221	GCA_900104805	
13R-gbRAST.gbk				4750 469	65.9	4598					

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 2022-02-18. 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 16178 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. 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 15 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 14 species clusters and the provided query strains were assigned to 1 of these. Moreover, user strains were located in 1 of 15 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 83.8 %. 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 94.2 %. The tree was rooted at the midpoint [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