TYPE (STRAIN) GENOME SERVER

JOB SUMMARY

PRINT DATE:	2022-02-19 00:10:31 +0100
JOB ID:	3d32e939-9e9e-430e-aa05-d4c971ad4f79
RESULT PAGE:	https://tygs.dsmz.de/user_results/show?guid=3d32e939-9e9e-430e-aa05-d4c971ad4f79

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
Azospirillum	potential new species		see [R1]

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
'Azospirillum'	*Azospirillum palustre* B2	82.3	[78.4 - 85.6]	65.2	[62.3 - 68.0]	81.9	[78.5 - 84.8]	0.03
'Azospirillum'	*Azospirillum melinis* TMCY 0552	81.0	[77.0 - 84.3]	62.0	[59.2 - 64.8]	80.0	[76.6 - 83.0]	0.1
'Azospirillum'	*Azospirillum lipoferum* VKM B-1519.	63.0	[59.3 - 66.6]	41.5	[39.0 - 44.1]	58.7	[55.5 - 61.9]	0.53
'Azospirillum'	*Azospirillum oryzae* COC8T	54.8	[51.3 - 58.3]	38.5	[36.0 - 41.0]	51.1	[48.0 - 54.2]	0.44
'Azospirillum'	*Azospirillum humicireducens* CCTCC AB 2012021	32.3	[28.9 - 35.9]	34.7	[32.3 - 37.2]	31.5	[28.6 - 34.6]	0.35
'Azospirillum'	*Azospirillum ramasamyi* KACC 14063	52.3	[48.8 - 55.7]	34.1	[31.6 - 36.6]	47.5	[44.5 - 50.6]	0.18
'Azospirillum'	*Azospirillum thiophilum* BV-s	52.9	[49.4 - 56.3]	31.4	[29.0 - 33.9]	46.9	[43.9 - 49.9]	0.35
'Azospirillum'	*Azospirillum picis* DSM 19922	39.7	[36.3 - 43.1]	28.3	[25.9 - 30.8]	36.0	[33.0 - 39.0]	0.95
'Azospirillum'	*Azospirillum agricola* CC-HIH038	33.6	[30.2 - 37.1]	25.7	[23.4 - 28.2]	30.6	[27.7 - 33.7]	1.94
'Azospirillum'	*Azospirillum doebereinerae* DSM 13131	33.1	[29.7 - 36.7]	25.4	[23.0 - 27.8]	30.2	[27.3 - 33.3]	1.08
'Azospirillum'	*Azospirillum doebereinerae* GSF71	33.0	[29.7 - 36.6]	25.4	[23.0 - 27.8]	30.2	[27.2 - 33.3]	1.08
'Azospirillum'	*Azospirillum griseum* L-25-5w-1	23.5	[20.2 - 27.2]	23.6	[21.3 - 26.1]	22.5	[19.7 - 25.6]	1.2
'Azospirillum'	*Azospirillum oleiclasticum* RWY-5-1-1T	18.1	[15.1 - 21.7]	21.2	[19.0 - 23.7]	17.8	[15.2 - 20.8]	1.75

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
Azospirillum oryzae COC8T	Xie and Yokota 2005	JCM 21588; NBRC 102291; CCTCC AB 204051; IAM 15130	*Azospirillum oryzae*	6749884	67.4	5859		PRJNA563036	SAMN12661881	GCA_008364795	
Azospirillum lipoferum VKM B-1519.	(Beijerinck 1925) Tarrand et al. 1979	LMG 13128; NRRL B-14654; NCIMB 11861; ATCC 29707; DSM 1691; JCM 1247; NBRC 102290; CIP 106280; NCAIM B.01801; sp. 59b	*Azospirillum lipoferum*; *Spirillum lipoferum*	7979458	67.3	6903		PRJNA563039	SAMN12661906	GCA_008364955	
Azospirillum oleiclasticum RWY-5-1-1T	Wu et al. 2020	KCTC 72259; CGMCC 1.13426T	*Azospirillum oleiclasticum*	7778031	69.6	7185		PRJNA224116	SAMN14851918	GCF_013423485	
Azospirillum picis DSM 19922	Lin et al. 2009	CCUG 55431; IMMIB TAR-3	*Azospirillum picis*	7011626	68.8	6303	Gp0538732				2928263341
Azospirillum palustre B2	Tikhonova et al. 2019	KCTC 62613; VKM B-3233	*Azospirillum palustre*	7989338	67.8	6821	Gp0253205	PRJNA414085	SAMN07776893	GCA_002573965	
Azospirillum ramasamyi KACC 14063	Anandham et al. 2019	NBRC 106460; M2T2B2	*Azospirillum ramasamyi*	6316263	68.0	5337	Gp0443122	PRJNA474386	SAMN09302594	GCA_003233655	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Azospirillum agricola CC-HIH038	Lin et al. 2016	BCRC 80909; JCM 30827	*Azospirillum agricola*	7597 657	69.7	6777	Gp0502503	PRJNA710668	SAMN18247811	GCA_017876095	
Azospirillum griseum L-25-5w-1	Yang et al. 2019	KCTC 62777; CGMCC 1.13672	*Azospirillum griseum*	5925 600	66.6	5164		PRJNA509942	SAMN10587723	GCA_003966125	
Azospirillum doebereinerae DSM 13131	Eckert et al. 2001	KCTC 12904; DSM 13131; GSF71	*Azospirillum doebereinerae*	6985 304	68.9	6248	Gp0401065	PRJNA546764	SAMN12025068		2828314363
Azospirillum humicireducens CCTCC AB 2012021	Zhou et al. 2013	KACC 16605; SgZ-5	*Azospirillum humicireducens*	3181 617	67.5	2833	Gp0150267	PRJNA318554	SAMN04858219	GCA_001639105	
Azospirillum thiophilum BV-s	Lavrinenko et al. 2010 emend. Hördt et al. 2020	DSM 21654; VKM B-2513	*Azospirillum thiophilum*	7609 458	68.2	6191	Gp0124194	PRJNA292868	SAMN03993951	GCA_001305595	
Azospirillum melinis TMCY 0552	Peng et al. 2006	LMG 23364; LMG 24250; DSM 17798; CGMCC 1.5340; CCBAU 5106001	*Azospirillum melinis*	7963 236	67.7	6920		PRJNA577426	SAMN13025581	GCA_013340935	
Azospirillum doebereinerae GSF71	Eckert et al. 2001	KCTC 12904; DSM 13131; GSF71	*Azospirillum doebereinerae*	6988 300	68.9	6031		PRJNA509943	SAMN10587691	GCA_003989665	
11R-gbRAST.gbk				7867 320	67.8	6903					

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 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 13 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 13 species clusters and the provided query strains were assigned to 1 of these. Moreover, user strains were located in 1 of 13 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 60.0 %. 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 96.0 %. 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