TYPE (STRAIN) GENOME SERVER

JOB SUMMARY

PRINT DATE:	2025-01-30 08:16:14 +0100
JOB ID:	d2f59ca6-2ac5-43ba-9312-1c0041883585
RESULT PAGE:	https://tygs.dsmz.de/user_results/show?guid=d2f59ca6-2ac5-43ba-9312-1c0041883585

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
'isolate_genome'	belongs to known species	*Ochrobactrum quorumnocens*	

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
'isolate_genome.fasta'	*Ochrobactrum gallinarum* Sa2BUA5	91.5	[88.4 - 93.8]	88.7	[86.2 - 90.7]	93.5	[91.2 - 95.2]	0.0
'isolate_genome.fasta'	*Ochrobactrum quorumnocens* A44 T	59.3	[55.7 - 62.9]	72.5	[69.5 - 75.3]	62.9	[59.6 - 66.1]	0.32
'isolate_genome.fasta'	*Brucella grignonensis* OgA9a	48.5	[45.1 - 51.9]	33.7	[31.2 - 36.2]	44.5	[41.5 - 47.5]	0.67
'isolate_genome.fasta'	*Brucella pituitosa* CCUG 50899	48.6	[45.2 - 52.0]	33.3	[30.9 - 35.8]	44.4	[41.4 - 47.5]	0.22
'isolate_genome.fasta'	*Brucella rhizosphaerae* PR17	57.2	[53.6 - 60.7]	32.9	[30.5 - 35.5]	50.7	[47.7 - 53.8]	0.47
'isolate_genome.fasta'	*Brucella pseudogrignonensis* CCUG 30717	53.3	[49.8 - 56.7]	27.5	[25.2 - 30.0]	45.3	[42.3 - 48.3]	0.51
'isolate_genome.fasta'	*Ochrobactrum chromiisoli* YY2X	30.1	[26.7 - 33.7]	25.5	[23.2 - 28.0]	28.0	[25.1 - 31.1]	0.51
'isolate_genome.fasta'	*Brucella thiophenivorans* DSM 7216	31.8	[28.5 - 35.4]	25.4	[23.1 - 27.9]	29.3	[26.3 - 32.4]	1.84
'isolate_genome.fasta'	*Brucella haematophila* CCUG 38531	19.5	[16.4 - 23.1]	22.8	[20.5 - 25.2]	19.1	[16.4 - 22.1]	3.19
'isolate_genome.fasta'	*Brucella pecoris* 08RB2639	21.0	[17.8 - 24.7]	22.8	[20.5 - 25.2]	20.4	[17.6 - 23.4]	2.47
'isolate_genome.fasta'	*Brucella lupini* LUP21	20.9	[17.7 - 24.5]	22.8	[20.5 - 25.2]	20.2	[17.5 - 23.3]	2.86
'isolate_genome.fasta'	*Brucella anthropi* ATCC 49188	21.3	[18.1 - 24.9]	22.7	[20.4 - 25.2]	20.6	[17.8 - 23.6]	2.64
'isolate_genome.fasta'	*Brucella tritici* LMG 18957	21.7	[18.4 - 25.3]	22.4	[20.1 - 24.8]	20.8	[18.0 - 23.9]	2.43

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
Brucella pituitosa CCUG 50899	(Huber et al. 2010) Hördt et al. 2020	DSM 22207	*Brucella pituitosa*; *Ochrobactrum pituitosum*	4275 050	53.7	8618		PRJNA445841	SAMN08800193	GCA_003049685	
Ochrobactrum gallinarum Sa2BUA5	Pallen 2024	DSM 112359; NCTC 14555	*Ochrobactrum gallinarum*	4977 758	53.5	4532		PRJNA543206	SAMN15803249	GCA_014836735	
Brucella pecoris 08RB2639	(Kämpfer et al. 2011) Hördt et al. 2020	08RB2639; CCUG 60088; DSM 23868; CCM 7822	*Brucella pecoris*; *Ochrobactrum pecoris*	5051 106	56.0	4730		PRJNA548045	SAMN11998268	GCA_006376675	
Brucella lupini LUP21	(Trujillo et al. 2006) Hördt et al. 2020	DSM 16930; NBRC 102587; LMG 22726	*Brucella lupini*; *Ochrobactrum lupini*	5582 811	56.3	5720	Gp0377653	PRJNA391246	SAMN07259926	GCA_002252535	
Brucella rhizosphaerae PR17	(Kämpfer et al. 2008) Hördt et al. 2020	CCUG 55411; DSM 19824; CCM 7493	*Brucella rhizosphaerae*; *Ochrobactrum rhizosphaerae*	4904 011	53.0	4928	Gp0372386	PRJNA391102	SAMN07258022	GCA_002252475	
Brucella pseudogrignonensis CCUG 30717	(Kämpfer et al. 2007) Hördt et al. 2020	CIP 109451; DSM 22354	*Brucella pseudogrignonensis*; *Ochrobactrum pseudogrignonense*	5531 557	54.0	5652	Gp0372383	PRJNA391102	SAMN07258024	GCA_002252525	
Brucella thiophenivorans DSM 7216	(Kämpfer et al. 2008) Hördt et al. 2020	CCUG 55412; CCM 7492	*Brucella thiophenivorans*; *Ochrobactrum thiophenivorans*	4364 781	51.6	3957	Gp0372384	PRJNA224116	SAMN07258021	GCF_002252445	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Brucella grignonensis OgA9a	(Lebuhn et al. 2000) Hördt et al. 2020	CCUG 46362; DSM 13338; NBRC 102586; LMG 18954	*Brucella grignonensis*; *Ochrobactrum grignonense*	4838 254	54.1	4564	Gp0372385	PRJNA224116	SAMN07258023	GCF_002252505	
Ochrobactrum quorumnocens A44 T	Krzyzanowska et al. 2024	LMG 30544; PCM 2957	*Ochrobactrum quorumnocens*	5645 291	53.2	5363		PRJNA224116	SAMN07259827	GCF_002278035	
Brucella anthropi ATCC 49188	(Holmes et al. 1988) Hördt et al. 2020	CIP 82.115; CCUG 24695; DSM 6882; JCM 21032; IFO 15819; NBRC 15819; NCTC 12168; LMG 3331	*Brucella anthropi*; *Ochrobactrum anthropi*	5205 777	56.1	4799	Gp0000090	PRJNA19485	SAMN02598421	GCA_000017405	640753038
Ochrobactrum chromiisoli YY2X	Yang et al. 2024	JCM 36000; CCTCC AB 2023035	*Ochrobactrum chromiisoli*	4650 646	53.0	4494		PRJNA224116	SAMN31656344	GCF_026241335	
Brucella tritici LMG 18957	(Lebuhn et al. 2000) Hördt et al. 2020	CCUG 47104; DSM 13340; NBRC 102585; SCII24	*Brucella tritici*; *Ochrobactrum tritici*	5183 744	55.9	4921		PRJNA573682	SAMN12821298	GCA_008932295	
Brucella haematophila CCUG 38531	(Kämpfer et al. 2007) Hördt et al. 2020	CIP 109452; DSM 22355	*Brucella haematophila*; *Ochrobactrum haematophilum*	5494 684	56.7	5029		PRJNA544772	SAMN11855631	GCA_005938105	
isolate_genome. fasta				5116 355	53.5	4770					

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 2025-01-29. 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 22308 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 4.0 [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 11 species clusters and the provided query strains were assigned to 1 of these. Moreover, user strains were located in 1 of 12 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 69.7 %. 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 94.5 %. 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