



PRINT DATE: 2023-04-11 11:58:50 +0200

JOB ID: 27bc6230-9952-4506-b2ee-740bed272f3d

RESULT PAGE: https://tygs.dsmz.de/user_results/show?guid=27bc6230-9952-4506-b2ee-740bed272f3d

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
'new_JS2_hybrid_assembly_contigs (1)'	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
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Fischerella sesquiti</i> JSC-11	56.2	[52.7 - 59.7]	45.3	[42.7 - 47.8]	54.4	[51.3 - 57.5]	0.41
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Pelatocladus maniniholoensis</i> HA4357-MV3	36.3	[32.9 - 39.8]	32.2	[29.8 - 34.7]	34.4	[31.4 - 37.4]	0.62
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Fischerella welwii</i> PCC 9431	35.2	[31.8 - 38.7]	31.3	[28.9 - 33.8]	33.3	[30.4 - 36.4]	0.45
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Brasilonema sennae</i> CENA114	13.4	[10.6 - 16.7]	21.8	[19.5 - 24.2]	13.7	[11.3 - 16.5]	1.86
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Mojavia pulchra</i> JT2-VF2	13.6	[10.9 - 17.0]	21.7	[19.5 - 24.2]	13.9	[11.5 - 16.7]	0.95
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Brasilonema octagenarum</i> UFV-E1	13.4	[10.6 - 16.7]	21.7	[19.5 - 24.2]	13.7	[11.3 - 16.5]	1.85
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Amazonocrinis nigriterrae</i> CENA67T	13.7	[10.9 - 17.0]	21.4	[19.2 - 23.8]	14.0	[11.6 - 16.8]	0.68
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Brasilonema bromeliae</i> SPC951	13.5	[10.8 - 16.9]	21.2	[18.9 - 23.6]	13.8	[11.4 - 16.6]	2.31
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Brasilonema angustatum</i> HA4187-MV1	13.4	[10.6 - 16.7]	21.1	[18.8 - 23.5]	13.7	[11.3 - 16.5]	1.77
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Iningainema tapete</i> BLCC-T55	13.2	[10.5 - 16.5]	21.0	[18.8 - 23.5]	13.5	[11.2 - 16.3]	0.8
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Desmonostoc vinosum</i> HA7617-LM4	13.4	[10.6 - 16.7]	21.0	[18.8 - 23.4]	13.7	[11.3 - 16.5]	0.87
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Atlanticothrix silvestris</i> CENA357 T	13.6	[10.8 - 16.9]	20.8	[18.6 - 23.2]	13.9	[11.5 - 16.7]	0.07
'new_JS2_hybrid_assembly_contigs (1).fa'	<i>Goleter apudmare</i> HA4340-LM2	13.5	[10.7 - 16.8]	20.7	[18.4 - 23.1]	13.8	[11.4 - 16.6]	1.53

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
<i>Fischerella welwii</i> PCC 9431	Walter et al. 2017		<i>Fischerella welwii</i>	7164 013	40.2	6017		PRJNA158821	SAMN02441755	GCA_000447295	
<i>Atlanticothrix silvestris</i> CENA357 T	Alvarenga et al. 2021	SP 428478	<i>Atlanticothrix silvestris</i>	6523 766	40.7	5501		PRJNA224116	SAMN16982919	GCF_016056295	
<i>Pelatocladus maniniholoensis</i> HA4357-MV3	Johansen and Vaccarino 2016	BISH 759444	<i>Pelatocladus maniniholoensis</i>	6704 691	40.0	5300		PRJNA727914	SAMN19065579	GCA_019359115	
<i>Goleter apudmare</i> HA4340-LM2	Miscoe and Johansen 2016	BISH 755086	<i>Goleter apudmare</i>	6711 310	42.2	5663		PRJNA727914	SAMN19065561	GCA_019359495	
<i>Fischerella sesquiti</i> JSC-11	Walter et al. 2017		<i>Fischerella sesquiti</i>	5378 468	41.0	4627		PRJNA61093	SAMN02256526	GCA_000231365	
<i>Amazonocrinis nigriterrae</i> CENA67T	Alvarenga et al. 2021	SP 428476	<i>Amazonocrinis nigriterrae</i>	8201 061	41.3	6899		PRJNA224116	SAMN16982918	GCF_016056305	
<i>Desmonostoc vinosum</i> HA7617-LM4	Miscoe and Johansen 2016	BISH 755083	<i>Desmonostoc vinosum</i>	7214 837	41.5	6082		PRJNA727914	SAMN19065558	GCA_019359515	
<i>Brasilonema angustatum</i> HA4187-MV1	Vaccarino and Johansen 2012	BISH 751683	<i>Brasilonema angustatum</i>	8509 522	42.4	6693		PRJNA727914	SAMN19065551	GCA_019359695	
<i>Mojavia pulchra</i> JT2-VF2	Řeháková and Johansen 2007	CCALA 691; BRY C37693	<i>Mojavia pulchra</i>	8192 881	41.6	6905		PRJNA727914	SAMN19065570	GCA_019359325	
<i>Brasilonema bromeliae</i> SPC951	Fiore et al. 2007	BRNM 1332	<i>Brasilonema bromeliae</i>	6238 926	43.0	5207		PRJNA477356	SAMN09468270	GCA_012912135	
<i>Brasilonema octagenarum</i> UFV-E1	Aguiar et al. 2008	BRNM HY1416	<i>Brasilonema octagenarum</i>	8258 280	42.5	6395		PRJNA477356	SAMN09467078	GCA_012912105	
<i>Iningainema tapete</i> BLCC-T55	Berthold et al. 2021	ULC 575; US 227647	<i>Iningainema tapete</i>	9491 544	41.4	8416		PRJNA662357	SAMN16082748	GCA_014698965	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Brasilonema sennae</i> CENA114	(Komárek 2003) Sant'Anna and Komárek 2011		<i>Brasilonema sennae</i> ; <i>Camptylonemopsis sennae</i>	8222 363	42.5	6371		PRJNA477356	SAMN09467079	GCA_006968745	
new_JS2_hybrid_ assembly_contigs (1).fa				7020 691	40.6	6060					

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 2023-04-11. 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 18853 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 93.5 %. 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 93.1 %. 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