



PRINT DATE: 2022-04-27 10:41:29 +0200

JOB ID: 0363512f-f1d6-4e53-b722-09ba6f748cbe

RESULT PAGE: https://tygs.dsmz.de/user_results/show?guid=0363512f-f1d6-4e53-b722-09ba6f748cbe

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
'STRPM-ROS4_final'	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
'STRPM-ROS4_final.fsa'	<i>Streptomyces albobogiseolus</i> JCM 4616	59.2	[55.6 - 62.8]	41.7	[39.2 - 44.2]	55.7	[52.6 - 58.9]	0.06
'STRPM-ROS4_final.fsa'	<i>Streptomyces viridodiataticus</i> JCM 4536	61.5	[57.8 - 65.1]	41.6	[39.1 - 44.2]	57.6	[54.4 - 60.7]	0.12
'STRPM-ROS4_final.fsa'	<i>Streptomyces variabilis</i> JCM 4422	67.1	[63.2 - 70.7]	40.7	[38.2 - 43.2]	61.6	[58.3 - 64.8]	0.05
'STRPM-ROS4_final.fsa'	<i>Streptomyces griseoincarnatus</i> JCM 4381	66.3	[62.5 - 70.0]	40.7	[38.2 - 43.2]	61.0	[57.8 - 64.2]	0.17
'STRPM-ROS4_final.fsa'	<i>Streptomyces althioticus</i> JCM 4344	68.0	[64.1 - 71.7]	40.6	[38.1 - 43.1]	62.3	[59.0 - 65.5]	0.1
'STRPM-ROS4_final.fsa'	<i>Streptomyces griseorubens</i> JCM 4383	68.0	[64.2 - 71.7]	40.5	[38.0 - 43.1]	62.3	[59.0 - 65.5]	0.02
'STRPM-ROS4_final.fsa'	<i>Streptomyces matensis</i> JCM 4277	69.7	[65.8 - 73.4]	40.5	[38.0 - 43.1]	63.6	[60.3 - 66.8]	0.22
'STRPM-ROS4_final.fsa'	<i>Streptomyces griseorubens</i> DSM 40160	68.7	[64.8 - 72.3]	40.5	[38.0 - 43.1]	62.8	[59.5 - 66.0]	0.09
'STRPM-ROS4_final.fsa'	<i>Streptomyces werraensis</i> JCM 4860	69.4	[65.5 - 73.1]	40.2	[37.7 - 42.8]	63.2	[59.9 - 66.4]	0.18
'STRPM-ROS4_final.fsa'	<i>Streptomyces gancidicus</i> JCM 4171	64.4	[60.6 - 68.0]	39.5	[37.0 - 42.0]	59.0	[55.8 - 62.1]	0.11
'STRPM-ROS4_final.fsa'	<i>Streptomyces griseoflavus</i> JCM 4479	47.1	[43.7 - 50.5]	32.5	[30.0 - 35.0]	43.0	[40.0 - 46.0]	0.14
'STRPM-ROS4_final.fsa'	<i>Streptomyces ambofaciens</i> ATCC 23877	34.6	[31.2 - 38.1]	26.8	[24.4 - 29.3]	31.7	[28.7 - 34.8]	0.22
'STRPM-ROS4_final.fsa'	<i>Streptomyces nigra</i> 452T	35.3	[31.9 - 38.8]	26.1	[23.8 - 28.6]	32.1	[29.1 - 35.1]	0.5
'STRPM-ROS4_final.fsa'	<i>Streptomyces longispororuber</i> JCM 4784	21.0	[17.8 - 24.6]	24.2	[21.9 - 26.7]	20.5	[17.7 - 23.5]	1.15

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>Streptomyces nigr</i> 452T	Chen et al. 2018	MCCC 1K03346; KCTC 39960	<i>Streptomyces nigr</i>	7641 029	71.9	6895	Gp0370472	PRJNA224116	SAMN08957528	GCF_003074055	
<i>Streptomyces griseorubens</i> DSM 40160	(Preobrazhenska ya et al. 1957) Pridham et al. 1958	BCRC 12104; CCRC 12104; NRRL B-3982; NRRL-ISP 5160; NCIMB 9846; CBS 505.68; ATCC 19767; ATCC 19909; DSM 40160; JCM 4383; IFO 12780; NBRC 12780; VKM Ac-1894; INA 6124/54; NCIB 9846; RIA 1047; UNIQEM 153	<i>Actinomyces griseorubens</i> ; <i>Streptomyces griseorubens</i>	7658 110	72.5	6883	Gp0505850				2913524612

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces ambofaciens</i> ATCC 23877	Pinnert-Sindico 1954 emend. Nouioui et al. 2018	BCRC 11857; BCRC (formerly CCRC) 11857; CCRC 11857; KCTC 9111; NRRL 2420; NRRL B-2516; NRRL ISP-5053; CBS 616.68; CECT 3101; DSM 40053; JCM 4204; JCM 4618; IFO 12836; IFO (now NBRC) 12836; NBRC 12836; RIA 1115	<i>Streptomyces ambofaciens</i>	8393 598	72.2	7793	Gp0002079	PRJNA93	SAMN03921868	GCA_001267885	
<i>Streptomyces matensis</i> JCM 4277	Margalith et al. 1959	NRRL B-2576; NRRL ISP-5188; NCIMB 9826; CBS 919.68; ATCC 23935; DSM 40188; JCM 4268; JCM 4651; IFO 12889; NBRC 12889; HAMBI 1048; IMET 42065; NCIB 9826; RIA 1142; RIA 570	<i>Streptomyces matensis</i>	7325 782	72.6	6731		PRJDB10510	SAMD00245408	GCA_014649175	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces griseoincarnatus</i> JCM 4381	(Preobrazhenskaya et al. 1957) Pridham et al. 1958 emend. Komaki 2021	9673/55; LMG 19316; BCRC 11481; CCRC 11481; BCRC 118481; NRRL B-5313; NRRL-ISP 5274; NCIMB 9825; CBS 838.68; ATCC 23623; ATCC 23917; DSM 40274; IFO 12871; NBRC 12871; AS 4.1409; CGMCC 4.1409; INA 9673; INA 9673/55; ISP 5274; NCIB 9825; RIA 1192	<i>Actinomyces griseoincarnatus</i> ; <i>Streptomyces griseoincarnatus</i>	7388 604	72.2	6759		PRJDB10510	SAMD00245431	GCA_014649575	
<i>Streptomyces werraensis</i> JCM 4860	Wallhäuser et al. 1964	BCRC 12038; CCRC 12038; NRRL B-5317; NRRL-ISP 5486; CBS 437.67; CBS 705.72; ATCC 14424; DSM 40486; DSM 40745; IFO 13404; NBRC 13404; RIA 1365	<i>Streptomyces werraensis</i>	7294 408	72.6	6499		PRJDB10510	SAMD00245533	GCA_014656175	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces albogriseolus</i> JCM 4616	Benedict et al. 1954 emend. Komaki 2021	40003; 7-A; BCRC 12230; CCRC 12230; KCTC 9675; NRRL B-1305; NRRL ISP-5003; NCIMB 9604; CBS 614.68; ATCC 23875; DSM 40003; JCM 4004; IFO 3413; IFO 3709; IFO 12834; NBRC 3413; NBRC 3709; NBRC 12834; VKM Ac-1200; CIP 104424; CIP 104428; HUT 6045; ISP 5003; NCIB 9604; RIA 1101	<i>Streptomyces albogriseolus</i>	8052 128	72.3	7414		PRJDB10510	SAMD00245488	GCA_014650475	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces variabilis</i> JCM 4422	(Preobrazhenskaya et al. 1957) Pridham et al. 1958	BCRC 11488; CCRC 11488; NRRL B-3984; NRRL-ISP 5179; CBS 568.68; ATCC 19815; ATCC 19930; DSM 40179; IFO 12825; NBRC 12825; VKM Ac-1854; IMET 42059; RIA 1095; UNIQEM 202	<i>Actinomyces variabilis</i> ; <i>Streptomyces variabilis</i>	7339 325	72.4	6699		PRJDB10510	SAMD00245451	GCA_014649935	
<i>Streptomyces griseoflavus</i> JCM 4479	(Krainsky 1914) Waksman and Henrici 1948	LMG 19344; BCRC 12232; CCRC 12232; NRRL B-5312; NRRL-ISP 5456; CBS 409.52; CBS 689.69; ATCC 25456; DSM 40456; IFO 13044; NBRC 13044; AS 4.1454; CGMCC 4.1454; VKM Ac-993; IMET 43530; RIA 1236	<i>Actinomyces griseoflavus</i> ; <i>Streptomyces griseoflavus</i>	7531 045	72.3	6910		PRJDB10510	SAMD00245456	GCA_014649995	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces gancidicus</i> JCM 4171	Suzuki 1957	IFM 1024; IFM 1024 (AAK-84); NRRL B-1872; NCIMB 12858; DSM 40935; IFO 15412; NBRC 15412	<i>Streptomyces gancidicus</i>	7265 211	72.3	6648		PRJDB10510	SAMD00245398	GCA_014649055	
<i>Streptomyces althioticus</i> JCM 4344	Yamaguchi et al. 1957 emend. Komaki 2021	BCRC 13686; CCRC 13686; KCTC 9752; NRRL B-3981; NRRL-ISP 5092; CBS 463.68; ATCC 19724; DSM 40092; IFO 12740; IFO 15956; NBRC 12740; NBRC 15956; AS 4.1608; CGMCC 4.1608; VKM Ac-705; ISP 5092; RIA 1005; UNIQEM 120	<i>Streptomyces althioticus</i>	7625 043	72.5	6991		PRJDB10510	SAMD00245418	GCA_014649355	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces griseorubens</i> JCM 4383	(Preobrazhenskaya et al. 1957) Pridham et al. 1958	BCRC 12104; CCRC 12104; NRRL B-3982; NRRL-ISP 5160; NCIMB 9846; CBS 505.68; ATCC 19767; ATCC 19909; DSM 40160; JCM 4383; IFO 12780; NBRC 12780; VKM Ac-1894; INA 6124/54; NCIB 9846; RIA 1047; UNIQEM 153	<i>Actinomyces griseorubens</i> ; <i>Streptomyces griseorubens</i>	7668 551	72.4	6981		PRJDB10510	SAMD00245433	GCA_014649615	
<i>Streptomyces viridodiastaticus</i> JCM 4536	(Baldacci et al. 1955) Pridham et al. 1958	BCRC 12458; CCRC 12458; NRRL B-5622; NRRL-ISP 5249; CBS 660.69; ATCC 25518; DSM 40249; IFO 13106; NBRC 13106; VKM Ac-1749; RIA 1298	<i>Actinomyces viridodiastaticus</i> ; <i>Streptomyces viridodiastaticus</i>	7674 156	72.5	7029		PRJDB10510	SAMD00245473	GCA_014655475	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
<i>Streptomyces longispororuber</i> JCM 4784	Waksman 1953	NRRL B-3736; NRRL B-5761; NRRL ISP-5599; NCIMB 9629; CBS 789.72; ATCC 27443; DSM 40599; IFO 13488; NBRC 13488; VKM Ac-1735; INA 11668/54; NCIB 9629; PCM 2396; RIA 1449	<i>Streptomyces longispororuber</i>	9000 117	73.6	7558		PRJDB10510	SAMD00245516	GCA_014656095	
STRPM-ROS4_final.fsa				7497 846	72.4	6786					

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 2022-04-27. 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 16600 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 14 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 10 species clusters and the provided query strains were assigned to 1 of these. Moreover, user strains were located in 1 of 10 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 61.5 %. 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 82.8 %. 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