TYPE (STRAIN) GENOME SERVER

JOB SUMMARY

PRINT DATE:	2024-08-27 18:19:09 +0200
JOB ID:	63774488-6e55-45b8-90a5-960cfb0d4a1b
RESULT PAGE:	https://tygs.dsmz.de/user_results/show?guid=63774488-6e55-45b8-90a5-960cfb0d4a1b

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
'GCF_004196355.1_ASM419635v1_genomic'	belongs to known species	*Serratia marcescens*	

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
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia nematodiphila* DSM 21420	89.2	[85.9 - 91.9]	85.4	[82.7 - 87.8]	91.3	[88.7 - 93.3]	0.1
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia marcescens* subsp. *sakuensis* KCTC 42172	85.7	[82.0 - 88.7]	73.2	[70.2 - 76.0]	86.4	[83.3 - 89.0]	0.02
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia marcescens* ATCC 13880	88.1	[84.6 - 90.8]	73.0	[70.0 - 75.8]	88.3	[85.4 - 90.7]	0.2
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia bockelmannii* S3 T	78.2	[74.3 - 81.7]	63.0	[60.1 - 65.8]	77.9	[74.5 - 81.0]	0.58
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia sarumanii* K-M0706	79.0	[75.0 - 82.5]	62.8	[59.9 - 65.6]	78.5	[75.1 - 81.6]	0.23
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia ureilytica* JCM 16474	79.1	[75.2 - 82.6]	62.6	[59.7 - 65.4]	78.6	[75.2 - 81.7]	0.17
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia montpellierensis* PsyLou2A	82.5	[78.7 - 85.8]	60.1	[57.2 - 62.8]	80.8	[77.4 - 83.8]	0.06
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia ficaria* NBRC 102596	60.2	[56.5 - 63.8]	34.7	[32.2 - 37.2]	53.7	[50.6 - 56.8]	0.28
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia entomophila* DSM 12358	63.0	[59.2 - 66.6]	33.7	[31.3 - 36.2]	55.3	[52.2 - 58.4]	0.71
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia plymuthica* NBRC 102599	48.9	[45.5 - 52.3]	28.0	[25.7 - 30.5]	42.5	[39.6 - 45.6]	3.68
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Serratia rubidaea* NBRC 103169	38.3	[34.9 - 41.8]	26.4	[24.0 - 28.9]	34.3	[31.4 - 37.4]	0.34
'GCF_004196355.1_ASM419635v1_genomic.fna'	*Klebsiella trevisanii* DSM 2688	15.8	[12.8 - 19.2]	20.5	[18.2 - 22.9]	15.8	[13.3 - 18.7]	4.48

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
Serratia entomophila DSM 12358	Grimont et al. 1988	A1; CIP 102919; ATCC 43705	*Serratia entomophila*	5147 822	58.9	4716		PRJNA727746	SAMN19031985	GCA_021462285	
Serratia marcescens subsp. *sakuensis* KCTC 42172	Ajithkumar et al. 2003	CIP 107489; DSM 17174; JCM 11315; KRED	*Serratia marcescens* subsp. *sakuensis*	5058 045	59.6	4719		PRJNA484649	SAMN09767465	GCA_003428265	
Serratia montpellierensis PsyLou2A	Blackburn et al. 2024	NRRL B-65689; LMG 32817; Lou2A	*Serratia montpellierensis*	5149 841	59.6	4734		PRJNA556397	SAMN12349642	GCA_020858655	
Serratia ureilytica JCM 16474	Bhadra et al. 2005	CCUG 50595; DSM 16952; LMG 22860; NIVA 51	*Serratia ureilytica*	5310 236	59.5	4988		PRJDB10510	SAMD00645019	GCA_039523675	
Serratia sarumanii K-M0706	Klages et al. 2024	DSM 116040; LMG 33111	*Serratia sarumanii*	5185 617	59.9	4769		PRJNA923618	SAMN32727621	GCA_029962605	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Serratia marcescens ATCC 13880	Bizio 1823	CFBP 4226; CIP 103235; NRRL B-2544; CCUG 1647; DSM 30121; JCM 1239; NBRC 102204; NCTC 10211; VKM B-1248; HAMBI 1286; LMG 2792	*Serratia marcescens*; *Serratia marcescens* subsp. *marcescens*	5092 451	59.8	4739	Gp0009486	PRJNA59561	SAMN02743269	GCA_000735445	
Serratia nematodiphila DSM 21420	Zhang et al. 2009	CGMCC 1.6853; KCTC 22130; DZ0503SBS1	*Serratia nematodiphila*	5224 920	59.5	4629	Gp0103438	PRJNA257492	SAMN02952129	GCA_000738675	
Serratia ficaria NBRC 102596	Grimont et al. 1981 emend. García-Fraile et al. 2020	4024; CIP 79.23; ATCC 33105; DSM 4569; JCM 1241; NCTC 12148; ICPB 4050; LMG 7881	*Serratia ficaria*	5261 721	59.9	4830	Gp0070998	PRJDB1514	SAMD00046907	GCA_001590885	
Serratia plymuthica NBRC 102599	(Lehmann and Neumann 1896) Breed et al. 1948	CIP 103239; ATCC 183; CCUG 14509; DSM 4540; JCM 1244; NCTC 12961; LMG 7886	*Bacterium plymuthicum*; *Serratia plymuthica*	5286 355	56.0	4829	Gp0023559	PRJDB268	SAMD00046908	GCA_001590925	

Strain	Authority	Other deposits	Synonyms	Base pairs	Percent G+C	No. proteins	Goldstamp	Bioproject accession	Biosample accession	Assembly accession	IMG OID
Serratia rubidaea NBRC 103169	(Stapp 1940) Ewing et al. 1973	CIP 103234; ATCC 27593; CCUG 10981; CCUG 9286; DSM 4480; JCM 1240; NCTC 12971; LMG 5019	*Bacterium rubidaeum*; *Serratia marinorubra*; *Serratia rubidaea*	4957 505	59.3	4492	Gp0023581	PRJDB269	SAMD00046903	GCA_001598675	
Klebsiella trevisanii DSM 2688	Ferragut et al. 1983	CIP 81.36; ATCC 33558; CUETM 78-120; Gavini K70; HAMBI 1303; LMG 3072	*Klebsiella trevisanii*	6215 359	55.2	5947	Gp0290555	PRJNA500331	SAMN10362885	GCA_004345285	2788499846
Serratia bockelmannii S3 T	Cho et al. 2020	DSM 110152; LMG 31535	*Serratia bockelmannii*	5284 737	59.1	4907		PRJNA554602	SAMN12274630	GCA_008011855	
GCF_004196355 .1_ASM419635v 1_genomic.fna				5183 522	59.6	4827					

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 2024-08-27. The TYGS analysis was subdivided into the following steps:

Determination of closely related type strains

The determination of closely related type strains did not succeed because not a single 16S rDNA gene sequence was detected in the provided user genomes. The subsequent analyses are thus only based on the provided genome data and the manually selected type strains, if any.

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 [3]. 100 distance replicates were calculated each. Digital DDH values and confidence intervals were calculated using the recommended settings of the GGDC 4.0 [2,3].

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 [4]. Branch support was inferred from 100 pseudo-bootstrap replicates each. The trees were rooted at the midpoint [5] and visualized with PhyD3 [6].

Type-based species and subspecies clustering

The type-based species clustering using a 70% dDDH radius around each of the 12 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 [7].

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 11 subspecies clusters.

Figure caption genome tree

Figure 2. Tree inferred with FastME 2.1.6.1 [4] 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 83.6 %. The tree was rooted at the midpoint [5].

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] 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

[4] 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

[5] Farris JS. Estimating phylogenetic trees from distance matrices. Am Nat. 1972;106: 645−667.

[6] 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

[7] 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