

**Complete genome sequence of *Thermosulfurimonas marina* SU872<sup>T</sup>, an anaerobic thermophilic chemolithoautotrophic bacterium isolated from a shallow marine hydrothermal vent.**

Maxime Allieux<sup>1</sup>, Mohamed Jebbar<sup>1</sup>, Galina Slobodkina<sup>2</sup>, Alexander Slobodkin<sup>2</sup>, Yann Moalic<sup>1</sup>, Anastasia Frolova<sup>2</sup>, Zongze Shao<sup>3</sup>, Karine Alain<sup>1\*</sup>

<sup>1</sup>Univ Brest, CNRS, IFREMER, LIA 1211 MicrobSea, Laboratoire de Microbiologie des Environnements Extrêmes LM2E, IUEM, Rue Dumont d'Urville, F-29280 Plouzané, France

<sup>2</sup>Winogradsky Institute of Microbiology, Research Center of Biotechnology of the Russian Academy of Sciences, Moscow, Russia

<sup>3</sup>Key Laboratory of Marine Genetic Resources, Third Institute of Oceanography, Ministry of Natural Resources, Xiamen 361005, China

**Supplementary Material S2: Text file providing missing information about comparative genomics, COG associations, other annotated genes of *Thermosulfurimonas marina* SU872<sup>T</sup> genome and associated gene locus tags.**

### S2.1. Comparative Genomics

*Thermosulfurimonas marina*'s 16S rRNA gene sequence was extracted from our reconstructed genome using Barrnap (v0.9 - <https://github.com/tseemann/barrnap>) and compared to the sequence published with the characterization of the strain (Frolova et al., 2018). The 16S rRNA gene sequence extracted with Barrnap resulted in a percentage identity of 99.67% with a partial 16S rRNA gene published previously (GenBank accession KY953157, Frolova et al., 2018). It

also shared 97.91% sequence identity with the 16S rRNA gene sequence of *Thermosulfurimonas dismutans* strain S95<sup>T</sup>, as published previously (Mardanov et al., 2016).

Average Nucleotide Identity (ANI) was calculated using the ANI calculator tool provided by the EzBioCloud web server (<https://www.ezbiocloud.net/tools/ani>) on genomes of the strain SU872<sup>T</sup> and *T. dismutans*, its closest relative organism (GenBank accession LWLG00000000, Mardanov et al., 2016) (Yoon et al., 2017). ANI score calculated against *T. dismutans* had an OrthoANIu value of 72.35 %, with an average aligned length of 810,696 bp. This value is much lower than the threshold criterion for prokaryotic species delineation proposed to be 95–96 % (Richter and Rosselló-Móra, 2009). DNA-DNA hybridization (DDH) estimate values were calculated using the genome-to-genome distance calculator (GGDC v2.1, formula 2) (Meier et al., 2013). The digital DNA-DNA hybridization estimate value between strain SU872<sup>T</sup> and *T. dismutans* S95<sup>T</sup> was 20.2%, which is far below the standard criterion (70%) for delineation of a prokaryotic species (Wayne et al., 1987), confirming that the strain SU872<sup>T</sup> belongs to a different species than *T. dismutans*.

## S2.2. COG associations

The major predicted COG categories (encompassing more than 2% of the CDSs) were related to translation-ribosomal structure-biogenesis (J) (8.4 %), energy production and conversion (C) (8.0 %), amino acid transport and metabolism (E) (7.5 %), replication-recombination-repair (L) (6.9 %), cell wall/membrane/envelope biogenesis (M) (6.5 %), signal transduction mechanisms (T) (5.2 %), coenzyme transport and metabolism (H) (5.0 %), posttranslational modification-protein turnover-chaperones (O) (4.8 %), cell motility (N) (4.3 %), inorganic ion transport and metabolism (P) (3.8%), intracellular trafficking-secretion-vesicular transport (U) (3.6 %), transcription (K) (3.4%), and carbohydrate transport and metabolism (G) (3.4 %), nucleotide transport and metabolism (F) (2.8%) and lipid transport and metabolism (I) (2.2%).

### S2.3. Other genes

No antibiotic resistance genes have been found with Mage platform, with CARD (v3.0.2) or RGI (v5.0.0) software. With BioCyc and Prokka, we found that *T. marina* possesses an arsenate reductase suggesting the capacity for utilization of arsenate as a terminal electron acceptor. However, in our laboratory experiments *T. marina* SU872<sup>T</sup> was not able to grow with arsenate (5 mM) and with either elemental sulfur (5 g/L) or molecular hydrogen as electron donors (data not shown).

### S2.4. Gene locus summary

Gene name	Gene associated locus (NCBI PGAP)
5-methyltetrahydrofolate corrinoid/iron sulfur protein methyltransferase	FVE67_RS00220
nitrogenase (molybdenum-iron type) associated sequences	FVE67_RS04245 ; FVE67_RS04250 ; FVE67_RS04255 ; FVE67_RS04260 ; FVE67_RS04275
Periplasmic Nap-type nitrate reductase	FVE67_RS02215
hydroxylamine reductase	FVE67_RS08695
hydroxylamine oxidase	FVE67_RS00020
glutamine synthetase	FVE67_RS00930
sulfate adenylyltransferase	FVE67_RS03090
adenylyl-sulfate reductase subunit A	FVE67_RS03075
adenylyl-sulfate reductase subunit B	FVE67_RS03080
manganese-dependent inorganic pyrophosphatase	FVE67_RS08780
dissimilatory sulfite reductase subunits alpha	FVE67_RS07950
dissimilatory sulfite reductase subunits beta	FVE67_RS07955
dissimilatory sulfite reductase subunits gamma	FVE67_RS07655
DsrM	FVE67_RS01460
DsrK	FVE67_RS01455
DsrJ	FVE67_RS01450
DsrO	FVE67_RS01445
DsrP	FVE67_RS01440
QmoA	FVE67_RS03070
QmoB	FVE67_RS03065
QmoC	FVE67_RS03060
dissimilatory sulfite reductase D	FVE67_RS07960
tetrathionate reductase subunit A associated sequences	FVE67_RS06180 ; FVE67_RS06900
tetrathionate reductase subunit B associated sequences	FVE67_RS02245 ; FVE67_RS02285 ; FVE67_RS06890 ; FVE67_RS08835
polysulfide reductase chain A	FVE67_RS02290
arsenate reductase	FVE67_RS06830

Table S2.4: Correspondences between the loci of the annotations by Prokka, Dfast, RAST and UniProtKB with the CDSs of the NCBI's automated prokaryotic genome annotation pipeline (PGAP). CDSs found with their associated loci, based on the assembly repository ASM1231758v1.

### Supplementary material associated references:

Mardanov, A.V., Beletsky, A.V., Kadnikov, V.V., Slobodkin, A.I., and Ravin, N.V., 2016. Genome analysis of *Thermosulfurimonas dismutans*, the first thermophilic sulfur-disproportionating bacterium of the phylum *Thermodesulfobacteria*. *Frontiers in Microbiology*, vol. 7, p. 950. <https://doi.org/10.3389/fmicb.2016.00950>

Meier-Kolthoff, J.P., Auch, A.F., Klenk, H.-P., Göker, M., 2013. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC Bioinformatics*, 14:60. <https://doi.org/10.1186/1471-2105-14-60>

Richter, M., Rosselló-Móra, R., 2009. Shifting the genomic gold standard for the prokaryotic species definition. *Proc. Natl. Acad. Sci. U. S. A.* 106 (45), 19126–19131. <https://doi.org/10.1073/pnas.0906412106>

Wayne, L.G., Brenner, D.J., Colwell, R.R., Grimont, P.A.D., Kandler, O. *et al.*, 1987. Report of the *ad hoc* committee on reconciliation of approaches to bacterial systematics. *International Journal of Systematic and Evolutionary Microbiology*, 37:463-464. <https://doi.org/10.1099/00207713-37-4-463>

Yoon, S. H., Ha, S. M., Lim, J. M., Kwon, S.J. & Chun, J., 2017. A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*, 110:1281–1286. <https://doi.org/10.1007/s10482-017-0844-4>