Genome sequence of *Sphingomonas sp. S17* isolated from an alkaline, hyperarsenic and hypersaline volcanic associated lake near 4000 meters above sea level in the Argentinean Puna

*Maria Eugenia Farias*, Santiago Revale, Estefania Mancini, Omar Ordoñez, Adrian Turjanski, Néstor Cortez and Martin P. Vazquez

1 Laboratorio de Investigaciones Microbiológicas de Lagunas Andinas (LIMLA), Planta Piloto de Procesos Industriales Microbiológicos (PROIMI), CCT-CONICET, Tucumán, Argentina. 2Instituto de Agrobiotecnología Rosario (INDEAR) – CONICET, Ocampo 210 bis, CCT-CONICET, Rosario, Argentina. 3Facultad de Ciencias Exactas y Naturales, Universidad de Buenos Aires. 4Instituto de Biología Molecular y Celular de Rosario (IBR-UNR-CONICET), Rosario, Argentina.

*Corresponding author. Mailing address: INDEAR-CONICET, Ocampo 210 bis, CCT-CONICET, Rosario, Argentina. Phone: (+54 341) 4448535 Fax: (+54 341) 4448557. Email: martin.vazquez@indear.com

**ABSTRACT**

The High-Altitude Andean Lakes (HAAL) in the Argentinian Puna-High Andes region represent an almost unexplored ecosystem exposed to extreme conditions (high UV irradiation, hypersalinity, drastic temperature changes, desiccation, high pH). Here we present the first genome sequence isolated from this extreme environment, a *Sphingomonas* sp.

**TEXT**

The *Sphingomonas sp. S17* was isolated from a modern stromatolite community in Socompa lake (3800 m altitude) placed near to the active volcano Socompa in Northwest Argentina. Environmental conditions of these lakes include a pH of 9, a salinity of 22% and an arsenic content of 32 mg/L (2, 3, 4, 5, 11).
The genome sequence was obtained using a Whole Genome Shotgun (WGS) strategy with a 454 GS Titanium pyrosequencer at INDEAR, Argentina. Assembly was done using 454 Newbler version 2.5.3 with the –urt option with a 63X genome coverage. This assembly generated 62 large contigs. The draft genome was 4,268,406 bases in length with a mean GC content of 65.74%.

Genome annotation was done using the Standard Operating Procedures (SOPs) for Prokaryotic annotation from ISGA (7) and from RAST annotation server (1). A total of 3892 Coding Sequences (CDS) and 52 structural RNAs (49 tRNAs) were predicted. Annotation covered 340 RAST subsystems (43%) with 1650 CDS, while 1251 CDS (32%) were classified as hypothetical proteins. The complete 16sRNA gene presented a maximum identity of 95% with *Sphingomonas wittichii* (10).

Consistent with the extreme environment and high UV irradiation, *Sphingomonas* sp. S17 presented a complete DNA repair system including a photolyase gene and SOS regulatory system with the umuCD operon for translesion synthesis. Neither of these CDS were found in the reference genome of *Sphingomonas wittichii RW1* (10). Accordingly, this bacterial genome also presented 24 genes devoted to the sulfur metabolism compared with only 4 genes in *S. wittichii*. A set of 95 genes was present in the subsystem for resistance to antibiotics and toxic compounds compared to only 41 in *S. wittichii*. The majority of these genes are devoted to the resistance of arsenic, chromium, fluoroquinolones and 18 genes for multidrug resistance efflux pumps. The genome also contained two copies of a NhaA type CDS for Na+/H+ antiporter (12) and the six subunits of the multi-subunit cation antiporter (Na+/H+), compatible with its alkaline and hypersaline environment.

As in other *Sphingomonas* genomes, a rich set of genes (47 CDS) was present in the subsystem for metabolism of aromatic compounds. It is interesting to note two other features absent from the *S. wittichii* genome: a) the presence of 5 CDS related to the prokaryotic immune system CRISPR (Cas1, Cas2, two Cas3 and Csn1) with seven associated Direct Repeats of 36 bases separated by spacers of 31 bases (8); b) the presence of the operon containing the LodA and LodB
CDS for the production of marinocine, a broad spectrum antibacterial protein (9). Finally, *Sphingomonas sp. S17* contains at least one megaplasmid of the incF type with a complete set of mobilization proteins (Variant 1 in the Conjugative transfer subsystem, RAST annotation) (6). The megaplasmid contains several resolvase/integrase proteins (8 CDS) and Transposases (5 CDS). Interestingly, it also harbors the umuCD operon. This represents the first genome sequence reported for the unexplored extreme HAAL ecosystem.

ACKNOWLEDGMENTS

The authors would like to thank other members of LIMLA for helpful discussion during the sequencing and analysis of this project. This work was supported by grants PICT-1707 and 1221, and PICT2007-0043 from ANPCyT-MINCyT, Argentina.

NUCLEOTIDE SEQUENCE ACCESSION NUMBERS

This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AFGG00000000. The version described in this paper is the first version, AFGG01000000.

REFERENCES


