

Draft Genome Sequence of *Tsukamurella* sp. Strain 1534

Lamia Oulmi,^{a,b} Aurore Gorlas,^a Grégory Gimenez,^a Catherine Robert,^a Abderrahmane Boulahrouf,^b Didier Raoult,^a and Véronique Roux^a

Aix Marseille Université, URMITE, UM63 CNRS 7278, IRD 198, INSERM 1095, Faculté de Médecine, Marseille, France,^a and Laboratoire Génie Microbiologique et Applications, Campus Chaâberss, Université Mentouri de Constantine, Constantine, Algérie^b

A draft genome sequence of *Tsukamurella* sp., an aerobic bacterium isolated from a human sputum specimen, is described here. A new virus or provirus, TPA4, was characterized.

The genus *Tsukamurella* includes Gram-positive rods isolated from sludge foams, arthropods, soil, and in opportunistic clinical infections. *Tsukamurella* sp. strain 1534 was isolated from a human sputum specimen. Recently, the whole genome of *Tsukamurella paurometabola* was published (9).

Both a shotgun sequencing and 3-kb paired-end sequencing were performed. The shotgun library was constructed with the GS Rapid library Prep kit (Roche). The library was constructed according to the 454 GS FLX Titanium paired-end protocol. The run was analyzed on the cluster through the GS Run-Viewer and Newbler assembler 2.5.3. A total of 482,748 paired-end filter wells were obtained from the both strategies and generated 168.2 Mb with an average length of 348.5 bp. The draft genome of *Tsukamurella* sp. strain 1534 (about 4.5 Mb) contains 4 scaffolds and 42 large contigs (>1,500 bp). One other contig had a size of 935 bp.

The G+C content was 71.1%. The genome contains 54 tRNA genes and encodes 4,400 putative proteins. Of the coding sequences, 89.4% could be assigned to Cluster of Orthologous Groups (COG) families. The sequences of the rRNAs were found in one contig, but its average coverage was 55×, compared with 37× for the average coverage of the whole genome. It was therefore possible to conclude that probably 2 ribosomal operons were present. Open reading frames (ORFs) were predicted using Prodigal (<http://prodigal.ornl.gov/>) with default parameters but the predicted ORFs were excluded if they were spanning a sequencing gap region. The predicted bacterial protein sequences were searched against the National Center for Biotechnology Information (NCBI) Nonredundant (NR), UNIPROT (7), and CharProtDB (6) databases using BLASTP and the Clusters of Orthologous Groups (COG) (11) database using RPSBLAST (1). The ARAGORN tool (5) was used to find tRNA genes, whereas ribosomal RNAs were found by using RNAmmer (4) and BLASTn against the NR database.

Reciprocal best BLAST analysis with the available genome of *Tsukamurella paurometabola* indicated that they shared 3,037 orthologs.

Bacteriophages were described in other *Tsukamurella* spp., and the PHAST server (12) and Prophage Finder (2) were used to identify potential proviruses in *Tsukamurella* sp. strain 1534 genome. The bacterium contains one genetic element of around 49.4 kb (with a G+C content of 70.1%) related to a virus named TPA4. A total of 59 open reading frames (ORFs) were recovered from TPA4, and 25 of them encode proteins sharing a high identity with proteins found in *Mycobacterium* viruses, especially the mycobac-

teriophage Giles (8). The preliminary annotation of TPA4 was performed, and the majority of the putative genes (a total of 35) encode hypothetical proteins. The ORFs with an attributed function (a total of 24) encode proteins involved in replication, such as DNA polymerase III, and recombination/integration and also those involved in phage morphogenesis. Despite the abundance of viruses in the biosphere (3), only two viruses have been isolated from *Tsukamurella* species: TPA1 and TPA2 (10). Further studies will be performed to determine the nature of TPA4 (provirus or virus), from which only 2 proteins share a high identity to TPA2 proteins. The chimeric architecture of TPA4 suggests the occurrence of gene transfer events during TPA4's evolution.

Nucleotide sequence accession numbers. The *Tsukamurella* sp. strain 1534 Whole Genome Shotgun project has been deposited in DDBJ/EMBL/GenBank under accession no. CAJY01000001 to CAJY01000042.

ACKNOWLEDGMENT

We are grateful to Laetitia Pizzo for technical assistance.

This research did not benefit from any external funding.

REFERENCES

- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J. Mol. Biol.* 215:403–410.
- Bose M, Barber RD. 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. *In Silico Biol.* 6:223–227.
- Hendrix RW. 2002. Bacteriophages: evolution of the majority. *Theor. Popul. Biol.* 61:471–480.
- Lagesen K, et al. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. *Nucleic Acids Res.* 35:3100–3108.
- Laslett D, Canback B. 2004. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. *Nucleic Acids Res.* 32:11–16.
- Madupu R, et al. 2012. CharProtDB: a database of experimentally characterized protein annotations. *Nucleic Acids Res.* 40(Database issue): D237–D241.
- Magrane M, Uniport Consortium. 29 March 2011. UniProt Knowledgebase: a hub of integrated protein data. *Database (Oxford)* 2011:bar009. doi:10.1093/database/bar009.

Received 17 July 2012 Accepted 27 July 2012

Address correspondence to Véronique Roux, veronique.roux@univ-amu.fr.

Copyright © 2012, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JB.01279-12

8. Morris P, Marinelli LJ, Jacobs-Sera D, Hendrix RW, Hatfull GF. 2008. Genomic characterization of mycobacteriophage Giles: evidence for phage acquisition of host DNA by illegitimate recombination. *J. Bacteriol.* **190**: 2172–2182.
9. Munk AC, et al. 2011. Complete genome sequence of *Tsukamurella paurometabola* type strain (no. 33T). *Stand. Genomic Sci.* **4**:342–351.
10. Petrovski S, Seviour RJ, Tillett D. 2011. Genome sequence and characterization of the *Tsukamurella* bacteriophage TPA2. *Appl. Environ. Microbiol.* **77**:1389–1398.
11. Tatusov RL, Galperin MY, Natale DA, Koonin EV. 2000. The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* **28**:33–36.
12. Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS. 2011. PHAST: a fast phage search tool. *Nucleic Acids Res.* **39**(Web server issue):W347–W352.