Genomics Data 5 (2015) 34–35

Contents lists available at ScienceDirect

Genomics Data

journal homepage: http://www.journals.elsevier.com/genomics-data/

Data in Brief

High-quality draft genome sequence of *Kocuria marina* SO9-6, an actinobacterium isolated from a copper mine

Daniel B.A. Castro [a,1], Letícia Bianca Pereira [a,1], Marcus Vinícius M. e Silva [c], Bárbara P. da Silva [b], Bruna Rafaella Z. Palermo [a], Camila Carlos [a], Daiane R.B. Belgini [b], Elmer Erasmo G. Limache [b], Gileno V. Jr Lacerda [b], Mariana B.P. Nery [b], Milene B. Gomes [b], Salatiel S. de Souza [d], Thiago M. da Silva [a], Viviane D. Rodrigues [a], Luciana C. Paulino [e], Renato Vicentini [a], Lúcio F.C. Ferraz [f], Laura M.M. Ottoboni [a,*]

[a] *Center for Molecular Biology and Genetic Engineering (CBMEG), State University of Campinas — UNICAMP, Campinas, Brazil*
[b] *Chemical, Biological and Agricultural Pluridisciplinary Research Center, State University of Campinas — UNICAMP, Campinas, Brazil*
[c] *Center APTA Citros Sylvio Moreira, Cordeirópolis, Brazil*
[d] *Department of Animal Biology, Institute of Biology, State University of Campinas — UNICAMP, Campinas, Brazil*
[e] *Center of Natural and Human Sciences, Federal University of ABC, Santo André, Brazil*
[f] *Department of Molecular Biology and Pharmacology, University of São Francisco, Bragança Paulista, Brazil*

ARTICLE INFO

Article history:
Received 31 March 2015
Received in revised form 6 May 2015
Accepted 10 May 2015
Available online 16 May 2015

Keywords:
Kocuria marina SO9-6
Genome
Copper mine
Aromatic compound degradation
Heavy metal tolerance

ABSTRACT

An actinobacterial strain, designated SO9-6, was isolated from a copper iron sulfide mineral. The organism is Gram-positive, facultatively anaerobic, and coccoid. Chemotaxonomic and phylogenetic properties were consistent with its classification in the genus *Kocuria*. Here, we report the first draft genome sequence of *Kocuria marina* SO9-6 under accession JROM00000000 (http://www.ncbi.nlm.nih.gov/nuccore/725823918), which provides insights for heavy metal bioremediation and production of compounds of biotechnological interest.

Specifications	
Organism	*Kocuria marina*
Strain	SO9-6
Sequencer or array type	Illumina HiSeq 2500
Data format	Processed
Experimental factors	Laboratory cultured strain
Experimental features	Very brief experimental description
Consent	High-quality draft genome of the actinobacterium *Kocuria marina* SO9-6
Sample source location	Brazilian Collection of Environmental and Industrial Microorganisms (CBMAI), accession number 1703

1. Direct link to deposited data

The draft genome sequence of *Kocuria marina* S09-6 has been deposited at DDBJ/EMBL/GenBank under the accession JROM00000000 (http://www.ncbi.nlm.nih.gov/nuccore/725823918), and this paper describes its first version.

2. Experimental design, materials and methods

The genus *Kocuria*, which belongs to the Micrococcaceae family, was first proposed by Stackebrandt et al. [1]. These bacteria were formerly classified in the genus *Micrococcus*, but were subsequently separated from it based on phylogenetic and chemotaxonomic analyses. The members of this genus are coccoid, Gram-positive, non-encapsulated, and aerobic, but *Kocuria kristinae* and *Kocuria marina* are exceptions, since the first is facultatively anaerobic and the second can grow in 5% CO_2 [1,2]. Members of the *Kocuria* genus have been isolated from different environments including marine sediment [2], saline desert soil [3] and fermented food [4]. Our strain was isolated from a sulfite ore containing partially oxidized chalcopyrite, obtained from the Sossego mine (6°25′45″S, 50°3′58″W) in Canaã dos Carajás, Brazil. To date, there is only one complete published genome of *Kocuria rhizophila* DC2201 (GenBank/EMBL/DDBJ accession number AP009152) [5] and

* Corresponding author. Tel.: +551935211099.
E-mail address: ottoboni@unicamp.br (L.M.M. Ottoboni).
[1] Both are first authors and equally contributed to this work.

http://dx.doi.org/10.1016/j.gdata.2015.05.006

Table 1
Kocuria marina SO9-6 genome statistics.

Attributes	Value
Genome size (bp)	3,066,141
Total contigs	62
GC content (%)	68.82
Protein-coding genes	2818
tRNA genes	48
rRNA genes	9
Genes assigned to subsystems	1260

four draft published genomes of *K. rhizophila* P7-4 (GenBank accession number AFID00000000) [6], *Kocuria atrinae* C3-8 (GenBank/EMBL/DDBJ accession number AJXN00000000) [7], *Kocuria* sp. strain UCD-OTCP (GenBank/EMBL/DDBJ accession number AOSQ00000000) [8], and *Kocuria palustris* (GenBank/EMBL/DDBJ accession number ANHZ00000000)[9]. Here, we report the draft genome of *K. marina* SO9-6, which is the first genome from this species to be sequenced.

Sequencing of the *K. marina* SO9-6 genome was performed at the Life Sciences Core Facility (LaCTAD) of the State University of Campinas (UNICAMP). A paired-end library (400 bp long) was sequenced using the HiSeq2500 system and yielded 154,561,434 paired-end reads of 100 bp. Reads were preprocessed with NGS QC Toolkit v.2.2.9 [10], with a quality cutoff of 20, minimum length of 70 bp, and removal of ambiguous bases. The genome size was estimated to be 3.24 Mbp based on k-mer count [11,12] and coverage of 4272×. The reads were subsampled to a genome coverage of approximately 250× and assembled with SPAdes [13]. IMAGE v.2.4 [14] was used to close the gaps, resulting in sequences with a total length of 3,066,141 bp in 62 contigs, N50 of 183,121 bp, and GC content of 68.82%. Open reading frames (ORFs) were predicted with GLIMMER-3 [15], and the Rapid Annotations using Subsystems Technology (RAST) server v.4.0 [16] was used for genome annotation. The *K. marina* SO9-6 genome sequence comprised 2818 coding sequences (CDSs), 48 tRNAs, and 9 rRNA genes (four 5S rRNA, three 23S rRNA, and two 16S rRNA). A total of 1260 (41.32%) protein-coding genes were classified in 373 subsystems, and a predicted function was assigned to 1208 of them (Table 1).

According to RAST, *K. rhizophila* DC2201 is the closest neighbor of our strain, encoding a type III polyketide synthase (T3pks) and a nonribosomal peptide synthetase, and is probably capable of degrading the aromatic compounds phenylacetate, protocatechuate, and homoprotocatechuate [5]. The functional annotation of *K. marina* SO9-6 revealed genes related to the degradation of aromatic compounds including the industrial water contaminant phenylacetic acid [17], benzoate (by means of hydroxylation), and phenylacetate. The antiSMASH v.2.0 [18] analysis revealed five known secondary metabolite clusters, one siderophore, two bacteriocins, one terpene, and one T3pks synthesis, in contigs KM0016, KM0022, KM0028, KM0045, and KM0055, respectively. A nearly complete route for butanol production from glycerol degradation was also identified, as well as genes involved in antibiotic resistance, tolerance to heavy metals (mercury, arsenic, zinc, and especially copper), and siderophore biosynthesis (implying the capacity to acquire iron), which could explain the organism's survival in the high metal content environment. The genes for aromatic compound degradation and heavy metal tolerance suggest that *K. marina* SO9-6 could be used to improve bioremediation processes in contaminated areas. Comparative genomic analyses of SO9-6 are in progress and will be published separately. This first genome of *K. marina* helps to provide insights into its survival in extreme environments and raises the possibility of genetic features that could be targeted in future bioremediation studies.

Acknowledgments

DBAC, LBP, MVMS, CC, GVLJ, MBPN, and TMS received fellowships from the São Paulo Research Foundation (FAPESP). BPS and EEGL received fellowships from the National Council for Technological and Scientific Development (CNPq). BRZP, SSZ, and VDR received fellowships from Coordination for the Improvement of Higher Education Personnel (CAPES). DRBB and MBG received fellowships from Petrobrás. The authors are grateful to Vale S.A. for the environmental mine sample.

References

[1] E. Stackebrandt, C. Koch, O. Gvozdiak, P. Schumann, Taxonomic dissection of the genus *Micrococcus*: *Kocuria* gen. nov., *Nesterenkonia* gen. nov., *Kytococcus* gen. nov., *Dermacoccus* gen. nov., and *Micrococcus* Cohn 1872 gen. emend. Int. J. Syst. Bacteriol. 45 (1995) 682–692.
[2] S.B. Kim, *Kocuria marina* sp. nov., a novel actinobacterium isolated from marine sediment. Int. J. Syst. Evol. Microbiol. 54 (2004) 1617–1620, http://dx.doi.org/10.1099/ijs.0.02742-0.
[3] W.J. Li, Y.Q. Zhang, P. Schumann, H.H. Chen, W.N. Hozzein, X.P. Tian, L.H. Xu, C.L. Jiang, *Kocuria aegyptia* sp. nov., a novel actinobacterium isolated from a saline, alkaline desert soil in Egypt. Int. J. Syst. Evol. Microbiol. 56 (2006) 733–737, http://dx.doi.org/10.1099/ijs.0.63876-0.
[4] E.J. Park, M.S. Kim, S.W. Roh, M.J. Jung, J.W. Bae, *Kocuria atrinae* sp. nov., isolated from traditional Korean fermented seafood. Int. J. Syst. Evol. Microbiol. 60 (2010) 914–918, http://dx.doi.org/10.1099/ijs.0.014506-0.
[5] H. Takarada, M. Sekine, H. Kosugi, Y. Matsuo, T. Fujisawa, S. Omata, E. Kishi, A. Shimizu, N. Tsukatani, S. Tanikawa, N. Fujita, S. Harayama, Complete genome sequence of the soil actinomycete *Kocuria rhizophila*. J. Bacteriol. 190 (2008) 4139–4146, http://dx.doi.org/10.1128/JB.01853-07.
[6] W.-J. Kim, Y.-O. Kim, D.-S. Kim, S.-H. Choi, D.-W. Kim, J.-S. Lee, H.J. Kong, B.-H. Nam, B.-S. Kim, S.-J. Lee, H.-S. Park, S.-H. Chae, Draft genome sequence of *Kocuria rhizophila* P7-4. J. Bacteriol. 193 (2011) 4286–4287, http://dx.doi.org/10.1128/JB.05334-11.
[7] Y.D. Nam, M.-J. Seo, S.I. Lim, S.L. Park, Genome sequence of *Kocuria atrinae* C3-8, isolated from jeotgal, a traditional Korean fermented seafood. J. Bacteriol. 194 (2012) 5996, http://dx.doi.org/10.1128/JB.01333-12.
[8] D.A. Coil, J.I. Doctor, J.M. Lang, A.E. Darling, J.A. Eisen, Draft genome sequence of *Kocuria* sp. strain UCD-OTCP (Phylum Actinobacteria). Genome Announc. 1 (2013) e00172-13, http://dx.doi.org/10.1128/genomeA.00172-13.
[9] G. Sharma, I. Khatri, S. Subramanian, Draft genome sequence of *Kocuria palustris* PEL. Genome Announc. 2 (2014) e01261-13, http://dx.doi.org/10.1128/genomeA.01261-13.
[10] R.K. Patel, M. Jain, NGS QC toolkit: a toolkit for quality control of next generation sequencing data. PLoS One 7 (2012) e30619, http://dx.doi.org/10.1371/journal.pone.0030619.
[11] G. Marçais, C. Kingsford, A fast, lock-free approach for efficient parallel counting of occurrences of k-mers. Bioinformatics 27 (2011) 764–770, http://dx.doi.org/10.1093/bioinformatics/btr011.
[12] R. Li, W. Fan, G. Tian, H. Zhu, L. He, J. Cai, Q. Huang, Q. Cai, B. Li, Y. Bai, Z. Zhang, Y. Zhang, W. Wang, J. Li, F. Wei, H. Li, M. Jian, J. Li, Z. Zhang, R. Nielsen, D. Li, W. Gu, Z. Yang, Z. Xuan, O.A. Ryder, F.C.-C. Leung, Y. Zhou, J. Cao, X. Sun, Y. Fu, X. Fang, X. Guo, B. Wang, R. Hou, F. Shen, B. Mu, P. Ni, R. Lin, W. Qian, G. Wang, C. Yu, W. Nie, J. Wang, Z. Wu, H. Liang, J. Min, Q. Wu, S. Cheng, J. Ruan, M. Wang, Z. Shi, M. Wen, B. Liu, X. Ren, H. Zheng, D. Dong, K. Cook, G. Shan, H. Zhang, C. Kosiol, X. Xie, Z. Lu, H. Zheng, Y. Li, C.C. Steiner, T.T.-Y. Lam, S. Lin, Q. Zhang, G. Li, J. Tian, T. Gong, H. Liu, D. Zhang, L. Fang, C. Ye, J. Zhang, W. Hu, A. Xu, Y. Ren, G. Zhang, M.W. Bruford, Q. Li, L. Ma, Y. Guo, N. An, Y. Hu, Y. Zheng, Y. Shi, Z. Li, Q. Liu, Y. Chen, J. Zhao, N. Qu, S. Zhao, F. Tian, X. Wang, H. Wang, L. Xu, X. Liu, T. Vinar, Y. Wang, T.-W. Lam, S.-M. Yiu, S. Liu, H. Zhang, D. Li, Y. Huang, X. Wang, G. Yang, Z. Jiang, J. Wang, N. Qin, L. Li, J. Li, L. Bolund, K. Kristiansen, G.K.-S. Wong, M. Olson, X. Zhang, S. Li, H. Yang, J. Wang, J. Wang, The sequence and de novo assembly of the giant panda genome. Nature 463 (2010) 311–317, http://dx.doi.org/10.1038/nature08696.
[13] A. Bankevich, S. Nurk, D. Antipov, A.A. Gurevich, M. Dvorkin, A.S. Kulikov, V.M. Lesin, S.I. Nikolenko, S. Pham, A.D. Prjibelski, A.V. Pyshkin, A.V. Sirotkin, N. Vyahhi, G. Tesler, M.A. Alekseyev, P.A. Pevzner, SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J. Comput. Biol. 19 (2012) 455–477, http://dx.doi.org/10.1089/cmb.2012.0021.
[14] I.J. Tsai, T.D. Otto, M. Berriman, Method improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps. Genome Biol. 11 (2010) R41, http://dx.doi.org/10.1186/gb-2010-11-4-r41.
[15] A.L. Delcher, D. Harmon, S. Kasif, O. White, S.L. Salzberg, Improved microbial gene identification with GLIMMER. Nucleic Acids Res. 27 (1999) 4636–4641, http://dx.doi.org/10.1093/nar/27.23.4636.
[16] R.K. Aziz, D. Bartels, A.A. Best, M. DeJongh, T. Disz, R.A. Edwards, K. Formsma, S. Gerdes, E.M. Glass, M. Kubal, F. Meyer, G.J. Olsen, R. Olson, A.L. Osterman, R.A. Overbeek, L.K. McNeil, D. Paarmann, T. Paczian, B. Parrello, G.D. Pusch, C. Reich, R. Stevens, O. Vassieva, V. Vonstein, A. Wilke, O. Zagnitko, The RAST server: rapid annotations using subsystems technology. BMC Genomics 9 (2008) 75, http://dx.doi.org/10.1186/1471-2164-9-75.
[17] L. Dsikowitzky, J. Schwarzbauer, Industrial organic contaminants: identification, toxicity and fate in the environment. Environ. Chem. Lett. 12 (2014) 371–386, http://dx.doi.org/10.1007/s10311-014-0467-1.
[18] K. Blin, M.H. Medema, D. Kazempour, M.A. Fischbach, R. Breitling, E. Takano, T. Weber, antiSMASH 2.0 — a versatile platform for genome mining of secondary metabolite producers. Nucleic Acids Res. 41 (2013) W204–W212, http://dx.doi.org/10.1093/nar/gkt449.