

Genome Sequence of *Streptomyces olindensis* DAUFPE 5622, Producer of the Antitumoral Anthracycline Cosmomycin D

Juan D. Rojas,^a Antonio Starcevic,^b Damir Baranašić,^b Maria A. Ferreira-Torres,^a Camilo A. Contreras,^a Leandro M. Garrido,^a Wellington L. Araújo,^a Robson F. de Souza,^a Jurica Zucko,^b Daslav Hranueli,^b Paul F. Long,^{c,d,e} John Cullum,^f Gabriel Padilla^a

Institute of Biomedical Sciences, University of São Paulo, São Paulo, Brazil^a; Faculty of Food Technology and Biotechnology, University of Zagreb, Zagreb, Croatia^b; Institute of Pharmaceutical Science^c and Department of Chemistry,^d King's College London, London, United Kingdom; Faculdade de Ciências Farmacêuticas, Universidade de São Paulo, São Paulo, Brazil^e; Department of Genetics, University of Kaiserslautern, Kaiserslautern, Germany^f

***Streptomyces olindensis* DAUFPE 5622, which was isolated from a Brazilian soil sample, produces the antitumor anthracycline cosmomycin D. The genome sequence is 9.4 Mb in length, with a G+C content of 71%. Thirty-four putative secondary metabolite biosynthetic gene clusters were identified, including the cosmomycin D cluster.**

Received 14 May 2014 Accepted 6 June 2014 Published 26 June 2014

Citation Rojas JD, Starcevic A, Baranašić D, Ferreira-Torres MA, Contreras CA, Garrido LM, Araújo WL, de Souza RF, Zucko J, Hranueli D, Long PF, Cullum J, Padilla G. 2014. Genome sequence of *Streptomyces olindensis* DAUFPE 5622, producer of the antitumor anthracycline cosmomycin D. *Genome Announc.* 2(3):e00541-14. doi:10.1128/genomeA.00541-14.

Copyright © 2014 Rojas et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 3.0 Unported license](https://creativecommons.org/licenses/by/3.0/).

Address correspondence to Gabriel Padilla, gpadilla@usp.br.

Streptomyces olindensis DAUFPE 5622 was isolated in the 1960s from a Brazilian soil sample. It produces the anthracycline cosmomycin D, which has antitumor activity and has attracted interest because of its distinctive glycosylation pattern (1–3). The first goal of the genome sequencing project was to determine the complete sequence of the cosmomycin biosynthetic cluster; the DNA sequence of part of the cluster was described earlier (1). A further goal was to identify additional secondary metabolite clusters in order to help isolate further interesting secondary metabolites from the strain. The genome sequences of *Streptomyces* strains often contain 20 to 30 potential secondary metabolite clusters, most of which do not correspond to known products of the strain (4). Although they have been called cryptic or silenced clusters, it is likely that they are expressed under appropriate physiological conditions. A data mining approach based on genome sequences is a promising route for isolating novel biologically active compounds.

The genome sequence was obtained using 454 pyrosequencing technology. The 650,799 reads obtained correspond to a coverage of about 29×. The assembly contains 120 contigs, with a total length of 9.4 Mb and a G+C content of 71%. The mean size of the contigs is 78 kb, and the N_{50} is 246 kb. The genome sequence should contain one linear plasmid of 76 kb and, if it is like other *Streptomyces* genomes, a single linear chromosome (5).

Glimmer (6) was used to determine potential protein-coding genes, which were annotated using the Rapid Annotations using Subsystems Technology (RAST) server (7). The genome was annotated using the NCBI Prokaryotic Genome Annotation Pipeline version 2.0 (8) (<http://www.ncbi.nlm.nih.gov/>). There were 8,287 predicted protein-coding genes and 4 rRNA gene clusters. Sixty-four tRNA sequences were predicted using the tRNAscan-SE Web server (9). Thirty-four secondary metabolite clusters were predicted using antiSMASH (10), including 7 polyketide synthases (PKS), 3 nonribosomal peptide synthetases (NRPS), 3 hybrid clusters (2 PKS-NRPS; 1 NRPS-terpene), 1 aminoglycoside, 3 sid-

erophores, 3 bacteriocins, 3 butyrolactones, 5 terpenes, 1 ectoine, and 5 clusters of other hypothetical functions. The cosmomycin cluster is about 40 kb long and contains 42 predicted genes, whose functions will be determined by further experimental studies. The sequence of the genome of *S. olindensis* will assist in the development of genetic engineering strategies for the production of new compounds with biotechnological potential.

Nucleotide sequence accession numbers. This whole-genome shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession no. [JJOH00000000](https://www.ncbi.nlm.nih.gov/nuclink/JJOH00000000). The version described in this paper is version [JJOH01000000](https://www.ncbi.nlm.nih.gov/nuclink/JJOH01000000).

ACKNOWLEDGMENTS

J.D.R., M.A.F.-T., and G.P. thank FAPESP grant (2009/52664-4), CAPES, and CNPq for financial support. This work was also supported by grant 09/5 (to D.H.) from the Croatian Science Foundation, Republic of Croatia, and by a cooperation grant of the German Academic Exchange Service (DAAD) and the Ministry of Science, Education and Sports, Republic of Croatia (to D.H. and J.C.). It was also supported by King's College London (to P.F.L.).

REFERENCES

1. Garrido LM, Lombó F, Baig I, Nur-e-Alam M, Furlan RL, Borda CC, Braña A, Méndez C, Salas JA, Rohr J, Padilla G. 2006. Insights in the glycosylation steps during biosynthesis of the antitumor anthracycline cosmomycin: characterization of two glycosyltransferase genes. *Appl. Microbiol. Biotechnol.* 73:122–131. <http://dx.doi.org/10.1007/s00253-006-0453-z>.
2. Furlan RL, Watt SJ, Garrido LM, Amarante-Mendes GP, Nur-e-Alam M, Rohr J, Braña A, Mendez C, Salas JA, Sheil MM, Beck JL, Padilla G. 2004. DNA-binding properties of cosmomycin D, an anthracycline with two trisaccharide chains. *J. Antibiot.* 57:647–654. <http://dx.doi.org/10.7164/antibiotics.57.647>.
3. Carvalho H, Garrido LM, Furlan RL, Padilla G, Agnoletto M, Guecheva T, Henriques JA, Saffi J, Menck CF. 2010. DNA damage induced by the anthracycline cosmomycin D in DNA repair-deficient cells. *Cancer Chemother. Pharmacol.* 65:989–994. <http://dx.doi.org/10.1007/s00280-010-1244-x>.
4. Walsh C, Fischbach M. 2010. Natural products version 2.0: connecting

- genes to molecules. *J. Am. Chem. Soc.* 132:2469–2493. doi: <http://dx.doi.org/10.1021/ja909118a>.
5. Chen CW, Yu TW, Lin YS, Kieser HM, Hopwood DA. 1993. The conjugative plasmid SLP2 of *Streptomyces lividans* is a 50 kb linear molecule. *Mol. Microbiol.* 7:925–932. <http://dx.doi.org/10.1111/j.1365-2958.1993.tb01183.x>.
 6. Delcher AL, Harmon D, Kasif S, White O, Salzberg SL. 1999. Improved microbial gene identification with Glimmer. *Nucleic Acids Res.* 27: 4636–4641. <http://dx.doi.org/10.1093/nar/27.23.4636>.
 7. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, Formsma K, Gerdes S, Glass EM, Kubal M, Meyer F, Olsen GJ, Olson R, Osterman AL, Overbeek RA, McNeil LK, Paarmann D, Paczian T, Parrelo B, Pusch GD, Reich C, Stevens R, Vassieva O, Vonstein V, Wilke A, Zagnitko O. 2008. The RAST server: Rapid Annotations using Subsystems Technology. *BMC Genomics* 9:75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
 8. Angiuoli SV, Gussman A, Klimke W, Cochrane G, Field D, Garrity G, Kodira CD, Kyrpides N, Madupu R, Markowitz V, Tatusova T, Thomson N, White O. 2008. Toward an online repository of Standard Operating Procedures (SOPs) for (meta)genomic annotation. *Omic* 12: 137–141. <http://dx.doi.org/10.1089/omi.2008.0017>.
 9. Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. *Nucleic Acids Res.* 33:W686–W689. <http://dx.doi.org/10.1093/nar/gki366>.
 10. Blin K, Medema MH, Kazempour D, Fischbach MA, Breitling R, Takano E, Weber T. 2013. antiSMASH 2.0—a versatile platform for genome mining of secondary metabolite producers. *Nucleic Acids Res.* 41: W204–W212. <http://dx.doi.org/10.1093/nar/gkt449>.