AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Whole-Genome Sequence of the Endophytic *Streptomyces* sp. Strain CBMAI 2042, Isolated from *Citrus sinensis*

Luciana Gonzaga de Oliveira,[a] Renata Sigrist,[a] Bruno Sachetto Paulo,[a] Markyian Samborskyy[b]

[a]Department of Organic Chemistry, University of Campinas, Campinas, São Paulo, Brazil
[b]Department of Biochemistry, University of Cambridge, Cambridge, England

ABSTRACT The whole-genome sequence of *Streptomyces* sp. strain CBMAI 2042, an endophytic actinobacterium isolated from *Citrus sinensis* branches, is described. The strain has the ability to inhibit the growth of *Xylella fastidiosa* and other human pathogens. *In silico* analysis highlighted the presence of nonribosomal peptide and polyketide synthases, revealing promising antibiotic assembly lines.

Streptomyces sp. strain CBMAI 2042 is an endophyte from *Citrus sinensis* capable of inhibiting the growth of *Bacillus megaterium*, *Staphylococcus aureus*, and *Candida albicans* pathogens (1). It is also known as a *Xylella fastidiosa* inhibitor, promoting the degradation of xantham gum (2) and other pathogens of *Citrus* subsp., such as *Geotrichum candidum* var. *citri-aurantii* and *Colletotrichum gloeosporioides*, representing an affordable alternative for sustainable agriculture practice (3). This actinobacterium was isolated from plant branch tissues. After surface sterilization, bark-free fragments (4 to 6 mm long) were plated onto petri dishes containing tryptic soy agar (TSA) supplemented with benomyl (50 μg ml^{-1}) for fungal growth inhibition. The plates were incubated at 28°C for 20 days (2). This endophyte was isolated from the cut pieces and identified based on the closest relative 16S rRNA matches (*Streptomyces anuatus*; the locus_tag prefix is STAN).

For high-molecular-weight genomic DNA (gDNA) extraction, a single colony was transferred to 50 ml of tryptic soy broth containing yeast extract (TSBY; Oxoid) and incubated at 28°C for 48 h. gDNA extraction followed the salting-out procedure (4), yielding nearly 3,000 ng μl^{-1}. The whole-genome draft was generated using Illumina shotgun TruSeq PCR-free and Nextera mate pair library prep using the manufacturer's protocol. Sequencing was carried out using V2 Illumina sequencing chemistry and run on a MiSeq instrument (2 $\times$ 250-bp paired-end [PE] sequencing).

The Illumina shotgun library produced 1.5 million reads in a total of 0.6 GB of data. The mate pair library produced 3.7 million reads in a total of 1 GB of data. The shotgun data were preprocessed without Illumina adapter sequences and used for bcl2fastq conversion. Reads were processed using an in-house Illumina adapter-trimming tool (fastq_miseq_trimmer) (5). The paired reads from the adapter-trimmed files were then preassembled using FLASh v1.2.11 (https://ccb.jhu.edu/software/FLASH/). The resulting combined read and not-combined read files were fed into Newbler v3.0 (Roche, 454 Life Sciences) for *de novo* assembly. The mate pair data set was preprocessed following the same workflow (fastq_miseq_trimmer tool), and the resulting split mate pair reads were assembled using Newbler v3.0. The assembly was then subjected to iterative Pilon polishing (v1.13), converted to Consed-compatible ace format, and checked using Consed (6, 7), generating 3 scaffolds, consisting of 53 contigs. This pipeline afforded 99% of the total genome coverage. The total genome size comprises 8,211 kbp (total length, 8,211,326 bp), with a G+C content of 68.4%.

Annotation was carried out using a customized pipeline based on FgeneSB as the

Citation Gonzaga de Oliveira L, Sigrist R, Sachetto Paulo B, Samborskyy M. 2019. Whole-genome sequence of the endophytic *Streptomyces* sp. strain CBMAI 2042, isolated from *Citrus sinensis*. Microbiol Resour Announc 8:e01426-18. https://doi.org/10.1128/MRA.01426-18.

Editor Jason E. Stajich, University of California, Riverside

Address correspondence to Luciana Gonzaga de Oliveira, lucianag@unicamp.br.

R.S. and B.S.P. contributed equally to this work.

Received 23 October 2018
Accepted 4 December 2018
Published 10 January 2019

open reading frame (ORF) predictor, operating in the *ab initio* mode (M. Samborskyy, unpublished data). The annotation results were edited using Artemis (8). Rapid Annotations using Subsystems Technology (RAST) server annotation (9) allowed for the identification of 7,198 candidate protein-coding genes. Additionally, 17 transcription factor initiators and subsystems related to the biosynthesis of secondary metabolites, fatty acids, lipids, isoprenoids, and siderophores were identified. Analysis through the antiSMASH 3.0.1 (10, 11) standalone platform highlighted 35 biosynthetic gene clusters (BGC). The production of valinomycin (NRP), alpiniamide (PK-NRP), and indigoidine (NRP) was confirmed by metabolite profile evaluation and heterologous host reconstitution (12, R. Sigrist, L. Gonzaga de Oliveira, 23 August 2017, Brazilian patent application BR1020170130239). This whole-genome sequencing reveals that the genome of endophyte *Streptomyces* sp. strain CBMAI 2042 encodes a large reservoir that qualifies for further biotechnological studies.

Data availability. This whole-genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession number RCOL00000000, BioProject number PRJNA487706, BioSample number SAMN09908177, and SRA number PRJNA487706 (SRA runs SRR8038340 to SRR8038346). The version described in this paper is the first version, RCOL01000000. *Streptomyces* sp. strain CBMAI 2042 was deposited at the Brazilian Collection of Environmental and Industrial Microorganisms (CBMAI).

ACKNOWLEDGMENTS

Illumina sequencing was performed at the Department of Biochemistry, University of Cambridge, Cambridge, England.

This sequencing work was supported by São Paulo Research Foundation grant 2014/12727-5. The authors also acknowledge the Support from CERSusChem-FAPESP-GSK (2014/50249-8). R.S. received a fellowship from São Paulo Research Foundation, FAPESP (2013/12598-8 and 2015/01013-4). B.S.P. and L.G.D.O. received a fellowship from National Council for Scientific and Technological Development, CNPq (140824/2017-0 and 313492/2017-4). L.G.D.O. is also grateful for the grant support provided by the program For Women in Science (2008, Brazilian Edition).

REFERENCES

1. Pedro LR, Giarola LR, Moraes S, Ellen D, Silva SG, Marcon J, João LA, Araujo WL, de Oliveira LG. 2015. Metabolic screening for PKS and NRPS in endophytic *Actinobacteria* from *Citrus reticulata*. Quim Nova 38:333–342. https://doi.org/10.5935/0100-4042.20150010.
2. Araújo WL, Marcon J, Maccheroni W, Jr, van Elsas JD, Van Vuurde JW, Azevedo JL. 2002. Diversity of endophytic bacterial populations and their interaction with *Xylella fastidiosa* in citrus plants. Appl Environ Microbiol 68:4906–4914. https://doi.org/10.1128/AEM.68.10.4906-4914.2002.
3. Singh R, Dubey AK. 2018. Diversity and applications of endophytic Actinobacteria of plants in special and other ecological niches. Front Microbiol 9:1767. https://doi.org/10.3389/fmicb.2018.01767.
4. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood DA. 2000. Practical Streptomyces genetics. The John Innes Foundation, Norwich, United Kingdom.
5. Ewing B, Hillier L, Wendl MC, Green P. 1998. Base-calling of automated sequencer traces using Phred. I. Accuracy assessment. Genome Res 8:175–185.
6. Gordon D, Green P. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics 29:2936–2937. https://doi.org/10.1093/bioinformatics/btt515.
7. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. Genome Res 8:195–202. https://doi.org/10.1101/gr.8.3.195.
8. Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B. 2000. Artemis: sequence visualization and annotation. Bioinformatics 16:944–945. https://doi.org/10.1093/bioinformatics/16.10.944.
9. 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, Parrello 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. https://doi.org/10.1186/1471-2164-9-75.
10. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. https://doi.org/10.1093/nar/gkv437.
11. 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. https://doi.org/10.1093/nar/gkt449.
12. Paulo BS, Sigrist R, Angolini CFF, Eberlin MN, de Oliveira LG. 2018. Gene deletion leads to improved valinomycin production by *Streptomyces* sp. CBMAI 2042. J Braz Chem Soc https://doi.org/10.21577/0103-5053.20180207.