AMERICAN SOCIETY FOR MICROBIOLOGY | genomeAnnouncements

Genome Sequence of *Streptomyces* sp. Strain RTd22, an Endophyte of the Mexican Sunflower

Fernanda O. Chagas,[a] Antonio C. Ruzzini,[b] Larissa V. Bacha,[a] Markyian Samborskyy,[c] Raphael Conti,[a]* Rita C. Pessotti,[a] Luciana G. de Oliveira,[d] Jon Clardy,[b] Monica T. Pupo[a]

School of Pharmaceutical Sciences of Ribeirão Preto, University of São Paulo, Ribeirão Preto, São Paulo, Brazil[a]; Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, Massachusetts, USA[b]; Department of Biochemistry, University of Cambridge, Cambridge, England[c]; Department of Organic Chemistry, University of Campinas, Campinas, São Paulo, Brazil[d]

* Present address: Raphael Conti, Department of Chemistry, Federal University of Espírito Santo, Vitória, Espírito Santo, Brazil.

F.O.C. and A.C.R. contributed equally to this work.

We report here the complete genome sequence of *Streptomyces* sp. strain RTd22, an endophytic actinobacterium that was isolated from the roots of the Mexican sunflower *Tithonia diversifolia*. The bacterium's 11.1-Mb linear chromosome is predicted to encode a large number of unknown natural products.

Received 25 May 2016 **Accepted** 30 May 2016 **Published** 21 July 2016

Citation Chagas FO, Ruzzini AC, Bacha LV, Samborskyy M, Conti R, Pessotti RC, de Oliveira LG, Clardy J, Pupo MT. 2016. Genome sequence of *Streptomyces* sp. strain RTd22, an endophyte of the Mexican sunflower. Genome Announc 4(4):e00693-16. doi:10.1128/genomeA.00693-16.

Address correspondence to Monica T. Pupo, mtpupo@fcfrp.usp.br.

Endophytic bacteria, in particular those in the genus *Streptomyces*, are potential reservoirs of new antibiotic and chemotherapeutic compounds. As part of a larger research program aimed at discovering biologically active small molecules from bacterial endophytes, we isolated *Streptomyces* sp. strain RTd22 from the Mexican sunflower, *Tithonia diversifolia* (*Asteraceae*). The plants were grown and harvested in Ribeirão Preto, São Paulo, Brazil, and the isolation and identification of endophytic bacteria were performed as previously described (1).

A draft genome sequence of *Streptomyces* sp. RTd22 was generated from paired-end libraries constructed with the Nextera library preparation kit. The sequencing (2 × 250-bp [300 cycle kit]) was performed using V2 Illumina sequencing chemistry run on a MiSeq instrument and produced 7.8 million reads. After filtering and adapter and quality trimming, the reads were assembled using the Lasergene SeqMan NGen version 3.1 (DNAStar) assembler program, converted to Consed-compatible ACE format, and checked using Consed (2, 3). The draft assembly consisted of 199 contigs (made into 97 scaffolds), with an average coverage of 70×. The 199 contigs comprise 11,180,448 bp, with a G+C content of 71.3%. Annotation was carried out using a customized pipeline based on FgenesB, operating in the *ab initio* mode, and the results were edited using Artemis (4). Genome annotation using RAST (5) predicted 9,577 coding sequences, including 93 RNA genes.

To overcome the assembly issues that are commonly associated with short-read data from high-G+C organisms, we resequenced RTd22 using PacBio single-molecule real-time (SMRT) sequencing technology (6). Specifically, an insert library of an 11.5 kb was prepared, and sequence data were generated from 4 SMRT cells run on a Pacific Biosciences RSII instrument using P6-C4 chemistry. *De novo* assembly was performed using the Hierarchical Genome Assembly Process (HGAP) (7), and after manual curation to remove low-quality bases, a single contiguous sequence of 11,142,275 bp with an average G+C content of 71.3% was produced. The RAST annotation server allowed the identification of 9,667 predicted protein-coding genes, including 88 RNA genes. Analysis of the 11.1-Mb RTd22 chromosome using antiSMASH (8) predicted ~40 biosynthetic gene clusters (BGCs) for secondary metabolite production. Among the 40 predicted BGCs, very few can be confidently assigned to a known natural product based on gene and nucleotide conservation. One exception is a BGC that is predicted to encode a himastatin-like molecule (9). More generally, the chromosome appears to be enriched in terpene-related genes with other abundant classes of BGCs, including nonribosomal peptides and several polyketide clusters that encode polyene macrolides. Altogether, genome sequencing has revealed that the endophyte *Streptomyces* sp. RTd22 encodes a large chemical reservoir that merits further studies.

Nucleotide sequence accession numbers. The whole-genome sequence has been deposited at DDBJ/ENA/GenBank under the accession numbers LXIC00000000 (Illumina) and CP015726 (PacBio). The version described in this paper is the first version.

ACKNOWLEDGMENTS

Illumina sequencing was performed at the Department of Biochemistry, University of Cambridge, Cambridge, England, and PacBio RSII sequencing was performed by the Duke Center for Genomic and Computational Biology Sequencing and Genomic Technologies Shared Resource core facility, Durham, NC.

We thank the Harvard Medical School Information Technology Department for access to and maintenance of the Orchestra High Performance Computing cluster.

FUNDING INFORMATION

This work, including the efforts of Monica T Pupo, was funded by São Paulo Research Foundation (FAPESP) (2013/07600-3, 2013/50954-0, and 2008/09540-0). This work, including the efforts of Luciana Gonzaga de Oliveira, was funded by São Paulo Research Foundation (FAPESP) (2011/06209-3). This work, including the efforts of Fernanda O Chagas, was funded by São Paulo Research Foundation (FAPESP) (2009/17695-6). This work, including the efforts of Larissa V Bacha, was funded by São Paulo Research Foundation (FAPESP) (2011/11703-7). This work, including the efforts of Raphael Conti, was funded by São Paulo Research Foundation (FAPESP) (2008/00812-7). This work, including the efforts of Rita C Pessotti, was funded by São Paulo Research Foundation (FAPESP) (2011/12910-6). This work, including the efforts of Jon Clardy, was funded by HHS | National Institutes of Health (NIH) (U19TW009872). This work, including the efforts of Fernanda O Chagas, was funded by MCTI | Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) (150572/2015-8). This work, including the efforts of Antonio Ruzzini, was funded by Gouvernement du Canada | Canadian Institutes of Health Research (CIHR) (201511MFE).

REFERENCES

1. **Conti R, Chagas FO, Caraballo-Rodriguez AM, Melo WGP, Nascimento AM, Cavalcanti BC, Moraes MO, Pessoa C, Costa-Lotufo LV, Krogh R, Andricopulo AD, Lopes NP, Pupo MT**. 2016. Endophytic actinobacteria from the Brazilian medicinal plant *Lychnophora ericoides* Mart. and the biological potential of their secondary metabolites. Chem Biodivers **13:**727–736. http://dx.doi.org/10.1002/cbdv.201500225.
2. **Gordon D, Green P**. 2013. Consed: a graphical editor for next-generation sequencing. Bioinformatics **29:**2936–2937. http://dx.doi.org/10.1093/bioinformatics/btt515.
3. **Gordon D, Abajian C, Green P**. 1998. Consed: a graphical tool for sequence finishing. Genome Res **8:**195–202. http://dx.doi.org/10.1101/gr.8.3.195.
4. **Rutherford K, Parkhill J, Crook J, Horsnell T, Rice P, Rajandream MA, Barrell B**. 2000. Artemis: sequence visualization and annotation. Bioinformatics **16:**944–945. http://dx.doi.org/10.1093/bioinformatics/16.10.944.
5. **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. http://dx.doi.org/10.1186/1471-2164-9-75.
6. **Eid J, Fehr A, Gray J, Luong K, Lyle J, Otto G, Peluso P, Rank D, Baybayan P, Bettman B, Bibillo A, Bjornson K, Chaudhuri B, Christians F, Cicero R, Clark S, Dalal R, Dewinter A, Dixon J, Foquet M, Gaertner A, Hardenbol P, Heiner C, Hester K, Holden D, Kearns G, Kong X, Kuse R, Lacroix Y, Lin S, Lundquist P, Ma C, Marks P, Maxham M, Murphy D, Park I, Pham T, Phillips M, Roy J, Sebra R, Shen G, Sorenson J, Tomaney A, Travers K, Trulson M, Vieceli J, Wegener J, Wu D, Yang A, Zaccarin D, Zhao P, et al.** 2009. Real-time DNA sequencing from single polymerase molecules. Science **323:**133–138. http://dx.doi.org/10.1126/science.1162986.
7. **Chin C-S, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J**. 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563–569. http://dx.doi.org/10.1038/nmeth.2474.
8. **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. http://dx.doi.org/10.1093/nar/gkv437.
9. **Ma J, Wang Z, Huang H, Luo M, Zuo D, Wang B, Sun A, Cheng Y-Q, Zhang C, Ju J**. 2011. Biosynthesis of himastatin: assembly line and characterization of three cytochrome P450 enzymes involved in the post-tailoring oxidative steps. Angew Chem Int Ed **50:**7797–7802. http://dx.doi.org/10.1002/anie.201102305.