

Draft genome of *Streptomyces* sp. AM6-12, a putative novel species for drug discovery isolated from Brazilian Restinga coastal soil

Matheus U. Oliveira,¹ Vicente Almeida Serafim da Silva,² Rafael Rocha Rangel,^{1,3} Nildo Alfredo Nhampossa,¹ João Ricardo Vidal Amaral,² Daniela Sales Alviano,³ Cinara Souza da Conceição,⁴ Carlos Alberto Xavier Gonçalves,⁴ Eamim Daidrê Squizani,⁴ Sheila da Silva,² Andrew Macrae,² Rodrigo Pires do Nascimento¹

AUTHOR AFFILIATIONS See affiliation list on p. 2.

ABSTRACT *Streptomyces* strain AM6-12 was isolated from Restinga de Marambaia soil and is able to completely inhibit the growth of *Escherichia coli*, *Staphylococcus aureus*, and *Pseudomonas aeruginosa*. Here we report on its genome sequence, with 7,958,853 bp and a 71.89% G+C content.

KEYWORDS antimicrobial activity, *Streptomyces*, bioactive compounds, secondary metabolites, actinomycetes, Kraken

Members of genus *Streptomyces* are gram-positive filamentous bacteria known for producing antimicrobial and biotechnologically relevant secondary metabolites (1). *Streptomyces* strain AM6-12 was isolated from Restinga de Marambaia soil (−23.068925°S, −43.960311°W), Rio de Janeiro, Brazil, an ecologically sensitive coastal area. Cultivation and characterization procedures previously performed (2) led to a new species description based on polyphasic analyses, under environmental sampling permit SISBIO/ICMBio N°34203-3. The strain exhibited antagonistic activity against *Staphylococcus aureus* MRSA BMB 9393, *Pseudomonas aeruginosa* ATCC 9027, and *Escherichia coli* ATCC 11229, on Mueller-Hinton agar at 35°C/48 h.

Genomic DNA extraction from spores followed a long-read sequencing-optimized protocol based on mechanical maceration, sodium dodecyl sulfate (SDS)/proteinase K lysis, and alcoholic precipitation (3). DNA was fragmented using G-Tubes (Covaris), and SMRTbell libraries were prepared using the SMRTbell Prep Kit 3.0 and Sequel II Binding Kit 3.2 (Pacific Biosciences), including AMPure PB size selection according to the manufacturer's instructions. Library concentration was measured using the DNA High Sensitivity Kit (Thermo Fisher Scientific) with a Qubit 4 fluorometer, and fragment size distribution was assessed using the Genomic DNA Kit (Agilent Technologies) with the TapeStation 4200 system. Sequencing was performed using the PacBio Sequel II platform (one SMRT Cell 8M; 2 h of pre-extension, 30 h of data acquisition), generating 60,669 raw reads with an N50 read length of 242,676 bp. Subreads quality control, error correction, and adapter trimming were conducted through the standard PacBio Circular Consensus Sequencing workflow prior to generating HiFi reads as input for genome assembly. Raw read quality was evaluated using FastQC v0.11.9 (4). *De novo* genome assembly was performed using Canu v2.2 (5) with default parameters. Final draft genome assembly was deposited in the NCBI, consisting of four contigs, totaling 7,958,853 bp. Polishing was performed using minimap2 2.28-r1209 and Racon v1.5 with default parameters (6, 7). QUAST v5.0.2 assembly (8), according to the method of Gurevich et al., reported an N50 of 341,342 bp, an L50 of four scaffolds, and a 71.89% G+C content. CheckM v1.1.3 (9) showed 100% completeness and 1.93% contamination. PROKKA v1.14.6 genome

Editor Silvia T. Cardona, University of Manitoba, Winnipeg, Manitoba, Canada

Address correspondence to Matheus U. Oliveira, matheuschoa@ufrj.br.

The authors declare no conflict of interest.

See the funding table on p. 2.

Received 8 January 2026

Accepted 7 March 2026

Published 25 March 2026

Copyright © 2026 Oliveira et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](#).

annotation (10) predicted 7,126 coding sequences, 46 rRNA genes, 1 tmRNA gene, and 91 tRNA genes.

Kraken2 v2.1.3 taxonomic assignment against the PLUS PFP 2023 database classified 98.37% of 68,114 reads within the genus *Streptomyces* (11). GTDB-Tk v2.4.0 genome-based classification confirmed genus-level placement, with the closest reference genome presenting 92.72% ANI ([GCF_030450125.1](https://doi.org/10.1093/gbe/ebaa012)), corroborating the previous genus-level classification (12). ANI calculations using OrthoANLu v3.8.3 and JSpeciesWS (13) yielded 91.01% identity to *Streptomyces malaysiensis* type strain MUSC 136. These values are below the accepted species threshold (95%–96%), indicating that strain AM6-12 likely represents a novel *Streptomyces* species.

Functional annotation by the Plabase pipeline (14) identified genes associated with nutrient acquisition and plant growth promotion, corroborated by HMMER/PFAM domain analysis. AntiSMASH v8.0 (15) detected multiple biosynthetic gene clusters, including similarities to ϵ -poly-L-lysine (51%), actinomycin-like clusters (45%), polyoxypeptin (39.7%), and pyrroloformamide (81%). Comparative analysis using clinker v0.0.28 (16) with default parameters demonstrated conserved cluster architecture relative to reference entries in the MIBiG database (17).

Streptomyces strain AM6-12 represents a putative novel species within the genus, with genomic features highlighting biotechnological and antimicrobial potential, warranting further investigation of its biosynthetic potential.

ACKNOWLEDGMENTS

A.M. and R.P.D.N. express their gratitude to Fundação Carlos Chagas Filho de Amparo à Pesquisa (FAPERJ) for financial support from the APQ1 grant. M.U.O. and N.A.N. thank Fundação Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) and the Postgraduate Research Program in Plant Biotechnology and Bioprocesses (PBV) for their grant support. R.R.R. and D.S.A. are grateful to the Conselho Nacional de Ciência e Tecnologia (CNPq) for grant funding.

All authors extend their gratitude to the Universidade Federal do Rio de Janeiro (UFRJ), FIRJAN, and SENAI/CETIQT for their contributions.

M.U.O. wrote the first draft; N.A.N., J.R.V.A., V.A.S.D.S., and S.D.S. performed genomics analyses and interpretation; R.R.R. performed biocontrol tests and analyses; C.S.D.C., E.D.S., and C.A.X.G. performed genome extraction and sequencing; S.D.S., A.M., and R.P.D.N. wrote the final text. All authors have approved the final article.

AUTHOR AFFILIATIONS

¹Universidade Federal do Rio de Janeiro, Escola de Química, Laboratório de Ecologia e Processos Microbianos, Rio de Janeiro, Brazil

²Universidade Federal do Rio de Janeiro, Instituto de Microbiologia, Laboratório de Biotecnologia Sustentável e Bioinformática Microbiana, Rio de Janeiro, Brazil

³Universidade Federal do Rio de Janeiro, Instituto de Microbiologia, Laboratório de Estrutura de Superfície de Microrganismos, Rio de Janeiro, Brazil

⁴Instituto SENAI de Inovação em Biossintéticos e Fibras, Rio de Janeiro, Brazil

AUTHOR ORCIDs

Matheus U. Oliveira  <http://orcid.org/0000-0003-2326-3623>

Sheila da Silva  <http://orcid.org/0000-0002-5020-4779>

FUNDING

Funder	Grant(s)	Author(s)
Fundação Carlos Chagas Filho de Amparo à Pesquisa do Estado do Rio de Janeiro		Andrew Macrae Rodrigo Pires do Nascimento

Funder	Grant(s)	Author(s)
Coordenação de Aperfeiçoamento de Pessoal de Nível Superior		Matheus U. Oliveira Nildo Alfredo Nhampossa
Conselho Nacional de Desenvolvimento Científico e Tecnológico		Rafael Rocha Rangel Daniela Sales Alviano

AUTHOR CONTRIBUTIONS

Matheus U. Oliveira, Writing – original draft | Vicente Almeida Serafim da Silva, Data curation, Investigation, Methodology, Software | Rafael Rocha Rangel, Formal analysis, Methodology, Validation | Nildo Alfredo Nhampossa, Data curation, Investigation, Methodology, Software | João Ricardo Vidal Amaral, Methodology, Software, Validation | Daniela Sales Alviano, Formal analysis, Supervision, Validation | Cinara Souza da Conceição, Data curation, Methodology, Software | Carlos Alberto Xavier Gonçalves, Data curation, Methodology, Software | Eamim Daidrê Squizani, Data curation, Methodology, Software | Sheila da Silva, Validation, Visualization, Writing – review and editing | Andrew Macrae, Conceptualization, Project administration, Supervision, Writing – review and editing | Rodrigo Pires do Nascimento, Supervision, Writing – review and editing

DATA AVAILABILITY

The genome sequence has been deposited at GenBank under accession number [JBIQWN000000000.1](https://www.ncbi.nlm.nih.gov/nuclseq/JBIQWN000000000.1), BioProject number [PRJNA1161424](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1161424), BioSample number [SAMN43777502](https://www.ncbi.nlm.nih.gov/biosample/SAMN43777502), and SRA number [SRX31707477](https://www.ncbi.nlm.nih.gov/sra/SRX31707477).

REFERENCES

- Hazarika S, Thakur D. 2020. Actinobacteria, p 443–476. In *Beneficial microbes in agro-ecology*. Academic Press.
- Rangel RR, Nhampossa NA, Dos Santos Costa D, Alviano DS, Alviano CS, do Nascimento RP. 2025. Antimicrobial potential of a new *Streptomyces* species isolated from a Brazilian tropical coastal shrubland. *Antonie Van Leeuwenhoek* 118:184. <https://doi.org/10.1007/s10482-025-02200-x>
- Alvarez-Arevalo M, Sterndorff EB, Faurdal D, Jørgensen TS, Mourched AS, Vuksanovic O, Saha S, Weber T. 2023. Extraction and Oxford Nanopore sequencing of genomic DNA from filamentous Actinobacteria. *STAR Protoc* 4:101955. <https://doi.org/10.1016/j.xpro.2022.101955>
- Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc/>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive *k*-mer weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>
- Heng L. 2018. Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics* 34:3094–3100. <https://doi.org/10.1093/bioinformatics/bty191>
- Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate *de novo* genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>
- Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
- Torsten S. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38:5315–5316. <https://doi.org/10.1093/bioinformatics/bt ac672>
- Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics* 32:929–931. <https://doi.org/10.1093/bioinformatics/btv681>
- Patz S, Gautam A, Becker M, Ruppel S, Rodríguez-Palenzuela P, Huson DH. 2021. PLABase: a comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. *bioRxiv*. <https://doi.org/10.1101/2021.12.13.472471>
- Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema MH, Weber T. 2021. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Res* 49:W29–W35. <https://doi.org/10.1093/nar/gkab335>
- Gilchrist CLM, Chooi Y-H. 2021. Clinker & clustermap.js: automatic generation of gene cluster comparison figures. *Bioinformatics* 37:2473–2475. <https://doi.org/10.1093/bioinformatics/btab007>
- Kautsar SA, Blin K, Shaw S, Navarro-Muñoz JC, Terlouw BR, van der Hoof JJJ, van Santen JA, Tracanna V, Suarez Duran HG, Pascal Andreu V, Selem-Mojica N, Alanjary M, Robinson SL, Lund G, Epstein SC, Sisto AC, Charkoudian LK, Collemare J, Lington RG, Weber T, Medema MH. 2020. MIBiG 2.0: a repository for biosynthetic gene clusters of known function. *Nucleic Acids Res* 48:D454–D458. <https://doi.org/10.1093/nar/gkz882>