



Data Article

Draft genome dataset of *Streptomyces griseoincarnatus* strain R-35 isolated from tidal pool sediments



Danielle Dana Mitchell, Jo-Marie Vreulink, Alaric Prins, Marilize Le Roes-Hill*

Applied Microbial and Health Biotechnology Institute, Cape Peninsula University of Technology, PO Box 1906, Bellville, Cape Town, 7530, South Africa

ARTICLE INFO

Article history:

Received 18 July 2024

Revised 23 August 2024

Accepted 29 August 2024

Available online 16 December 2024

Dataset link: Genomics dataset of marine isolate *Streptomyces griseoincarnatus* strain R-35 (Original data)

Dataset link: Multi-omics data from marine sites along the Western Cape, South Africa (Original data)

Keywords:

Actinobacteria

Genomics

Marine

Streptomyces

Tidal pool

ABSTRACT

The marine isolate, *Streptomyces griseoincarnatus* strain R-35, was isolated from marine sediments collected from the Glencairn Tidal Pool, Table Mountain National Park, Cape Town, South Africa. The genomic DNA was sequenced using the Ion Torrent GeneStudio™ S5 platform, and the *de novo* assembly was performed using the SPAdes assembler on the Centre for High Performance Computing (CHPC) Lengau Cluster located at the CSIR, Rosebank, South Africa. The draft genome assembly consisted of 722 contigs totaling 7,625,174 base pairs and a G+C% content of 72.2 mol%. Genome completeness and genome contamination were determined as 99.12% and 0.92%, respectively. Genome annotations performed using the Rapid Annotation with Subsystem Technology (RAST) and the Bacterial and Viral Bioinformatics Resource Centre (BV-BRC) determined the presence of 7996 coding sequences (CDS), 63 transfer RNAs (tRNAs), and six ribosomal RNAs (rRNAs). A total of 2570 hypothetical proteins were assigned, and 5246 proteins were assigned to function. The phylogenomic positioning of *S. griseoincarnatus* strain R-35 was determined using the Type Strain Genome Server (TYGS) and was found to be related to *S. griseoincarnatus* JCM 4381^T, with a digital DNA-DNA hybridisation

* Corresponding author.

E-mail address: leroesm@cput.ac.za (M. Le Roes-Hill).

Social media:  @PrinsAlaric (A. Prins)

<https://doi.org/10.1016/j.dib.2024.111235>

2352-3409/© 2024 Published by Elsevier Inc. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

(dDDH) value of 84.1%, and an OrthoANlu value of 98.22%. The CARD RGI algorithm on Proksee predicted the presence of 6,107 antimicrobial resistance (AMR) features, 27 biosynthetic gene clusters (BGCs) were predicted using antiSMASH, while 189 carbohydrate-active enzymes (CAZymes) were predicted using dbCAN3. The raw genome sequencing data has been submitted to the National Center for Biotechnology (NCBI) under the BioProject ID PRJNA1129156 (BioSample ID Accession Number: SAMN42145163; Short Read Archive (SRA) Accession: SRR29633055; <https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1129156>).

© 2024 Published by Elsevier Inc.
This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>)

Specifications Table

Subject	Microbial Genomics
Specific subject area	Microbiology, Genomics, Biotechnology
Type of data	Raw and processed data (tables, graphs, figures)
Data collection	<i>Streptomyces griseoincarnatus</i> strain R-35 was isolated from marine sediments collected from the Glencairn tidal pool, Table Mountain National Park. Cultures were maintained on ISP2 prepared with artificial seawater. Genomic DNA was extracted from the strain using a ZymoResearch QuickDNA Fecal/Soil Microbe Prep Kit. The DNA library was prepped using the Ion 540™ Chef Kit and sequenced using an Ion GeneStudio™ S5 Prime system. A draft genome assembly was performed using SPAdes on the Centre for High Performance Computing Lengau Cluster, CSIR, Rosebank, South Africa. Downstream processing was performed using Quast, CheckM, RAST, BV-BRC, TYGS, ANI calculator, Proksee, and antiSMASH.
Data source location	The isolate, <i>Streptomyces griseoincarnatus</i> strain R-35 was isolated from marine sediment collected from the Glencairn Tidal Pool (co-ordinates -34.162481, 18.432022). The isolate is maintained in the culture collection at the Applied Microbial and Health Biotechnology Institute (AMHBI), Cape Peninsula University of Technology, Cape Town, South Africa.
Data accessibility	Repository name: The raw genome sequence reads for <i>Streptomyces griseoincarnatus</i> strain R-35 is deposited and publicly available on NCBI under BioProject PRJNA1129156 with the BioSample accession SAMN42145163. Data identification number: The short read archive (SRA) is available on NCBI under accession number SRR29633055. Direct URL to data: https://www.ncbi.nlm.nih.gov/bioproject/PRJNA1129156 Secondary Data: Outputs from the bioinformatics tools applied are publicly available on the Mendeley Data Repository at the following link: 10.17632/4tykmn5gyp.3
Related research article	Not applicable

1. Value of the Data

- The dataset details a core genome analysis describing a marine actinobacterial isolate, *Streptomyces griseoincarnatus* strain R-35, isolated from a marine site not previously explored for actinobacteria.
- The dataset provides a potential genomic resource for exploring biosynthetic genes, including those encoding for antimicrobials and enzymes.
- The dataset is relevant to clinical and industrial biotechnology applications.
- The dataset expands our current taxonomic understanding of strains identified as *Streptomyces griseoincarnatus* and their geographical distribution.

2. Background

The genus *Streptomyces* is a large, phenotypically diverse bacterial genus in the phylum *Actinomycetota* and currently contains 730 validly published child taxa [1]. With some exceptions, typical morphological characteristics include extensive branching substrate and aerial mycelia [2]. These organisms typically have large, linear genomes ranging from 8.5 to 12Mbp in size and high G+C content of between 67 and 78% [3]. *Streptomyces* species are abundant in various soil environments where they play a key role in the biogeochemical cycle but have also been extensively described in marine environments as free-living organisms or as symbionts [4]. *Streptomyces* species are prolific producers of secondary metabolites, accounting for about 17% of all known active secondary metabolites [5]. These secondary metabolites exhibit various properties, including antimicrobial, anti-inflammatory, antiviral, antitumour, and anti-diabetic activities [6], largely attributed to many biosynthetic gene clusters encoded for in their large genomes [7]. As part of a larger coastal biodiversity survey along the Western Cape, South Africa, *Streptomyces griseoincarnatus* strain R-35 was isolated from marine sediments collected from the Glencairn tidal pool, Table Mountain National Park and explored for its biosynthetic potential.

3. Data Description

The raw genome sequence data is deposited on NCBI under BioProject at accession PRJNA1129156.

At the root of the secondary data (analyses outputs) folder, a summary document of file and folder structure can be found (file: *R-35_Analyses_DirectoryContents.txt*)

The draft genome assembly of *Streptomyces griseoincarnatus* strain R-35 (file: *001_Assembly > R-35_gDNA_assembly.fasta*) contains 722 contigs, with the largest contig being 157,132bp in size. The draft genome is 7,625,174bp in length with a G+C content of 72.2 mol%, and an N₅₀ score of 25 615bp (file: *002_Quast > R-35_QUAST.html*). The completeness and contamination of the genome of strain R-35 were 99.12% and 0.92%, respectively (Table 1, file: *003_CheckM > tab_text > CheckM_summary_table.tsv*).

Genome annotations with the Rapid Annotation with Subsystem Technology (RAST) tool and the Bacterial and Viral Bioinformatics Resource Centre (BV-BRC) determined the presence of 7,996 coding sequences (CDS), 63 transfer RNA (tRNA) genes, and six ribosomal RNA (rRNA) genes (full RAST feature list file: *004_Annotation > R-35_RAST.xls/gbk*). A total of 312 subsystem features were annotated, contributing 18% of the detected features (Fig. 1).

The BV-BRC annotation included 2570 hypothetical proteins and 5426 proteins assigned to function (Table 2). Of these, 1148 proteins were assigned with Enzyme Commission (EC) numbers, 1213 with Gene Ontology (GO) assignments, and 1113 with KEGG pathway assignments. The BV-BRC annotation also includes two protein families – genus-specific protein families (PLFams) and cross-genus protein families (PGFams), of which there were 7307 and 7402, respectively (file: *004_Annotations > R-35_FullGenomeReport_PATRIC.html*).

The phylogenomic positioning of *S. griseoincarnatus* strain R-35 was determined using the Type Strain Genome Server (TYGS) (Fig. 2). Strain R-35 is closely related to *Streptomyces ery-*

Table 1
Summary of draft genome assembly metrics for *Streptomyces griseoincarnatus* strain R-35.

Genome length (bp)	7,625,174
Number of contigs	722
G+C mol%	72.2
N ₅₀ (bp)	25,615
Marker lineage (CheckM)	f__Streptomycetaceae
Completeness (%)	99.12
Contamination (%)	0.92

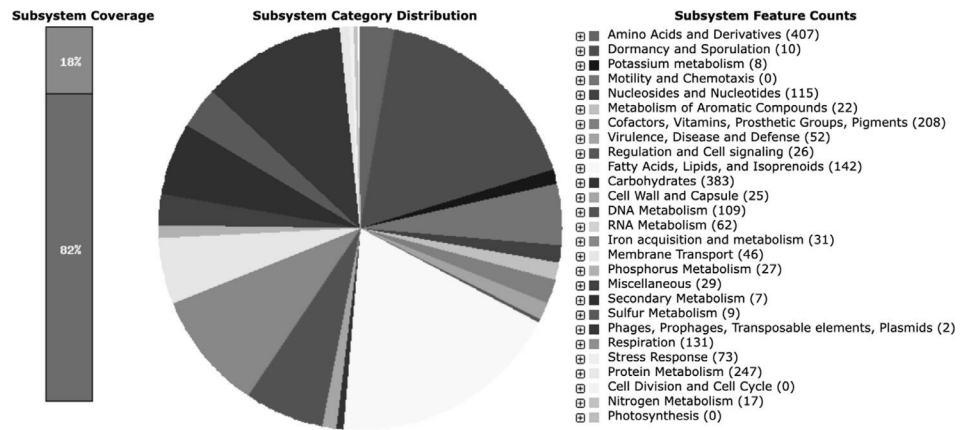


Fig. 1. Subsystem of the draft genome of *Streptomyces griseoincarnatus* strain R-35 annotated through the RAST Annotation Server.

Table 2
Protein features annotated in BV-BRC for draft genome assembly of *Streptomyces griseoincarnatus* strain R-35.

Feature	Total
Hypothetical Proteins	2,570
Proteins with functional assignments	5,426
Proteins with EC number assignments	1,418
Proteins with GO assignments	1,213
Proteins with KEGG pathways assignments	1,113
Proteins with PATRIC genus-specific family (PLFam) assignments	7,307
Proteins with PATRIC cross-genus family (PGFam) assignments	7,402

Table 3
Summary of specialty genes present in the genome of *Streptomyces griseoincarnatus* strain R-35 as predicted using BV-BRC.

Type of Gene	Source Database	Amount
Antibiotic Resistance	CARD	6
Antibiotic Resistance	NDARO	1
Antibiotic Resistance	PATRIC	52
Drug Target	DrugBank	10
Drug Target	TTD	1
Transporter	TCDB	61
Virulence Factor	PATRIC_VF	8
Virulence Factor	Victors	2

throgrius JCM 9650^T, *Streptomyces labedae* JCM 9381^T, *Streptomyces griseoincarnatus* JCM 4381^T, and *Streptomyces variabilis* JCM 4422^T. A digital DNA-DNA hybridization (dDDH) value of 84.1% and a minimal G+C % difference of 0.3 assigned strain R-35 as *S. griseoincarnatus* (file: 005_Phylogenomics > R-35_TYGS_JobResults.pdf). As confirmed by the OrthoANlu value (98.2%) against *S. griseoincarnatus* JCM 4381^T (file: 005_Phylogenomics > R-35_OrthoANlu_results.pdf).

Several specialty genes were annotated in BV-BRC with homology to known transporters, virulence factors, drug targets, and antibiotic resistance genes. Table 3 summarises homologous gene searches against several databases, including CARD, NDARO, DrugBank, and Victors.

Furthermore, antimicrobial resistance (AMR) genes were searched for using Proksee (CARD RGI) by including partial hits to compensate for shorter contigs in the genome assembly. A total of 6,107 AMR features were detected (file: 004_Annotations > R-

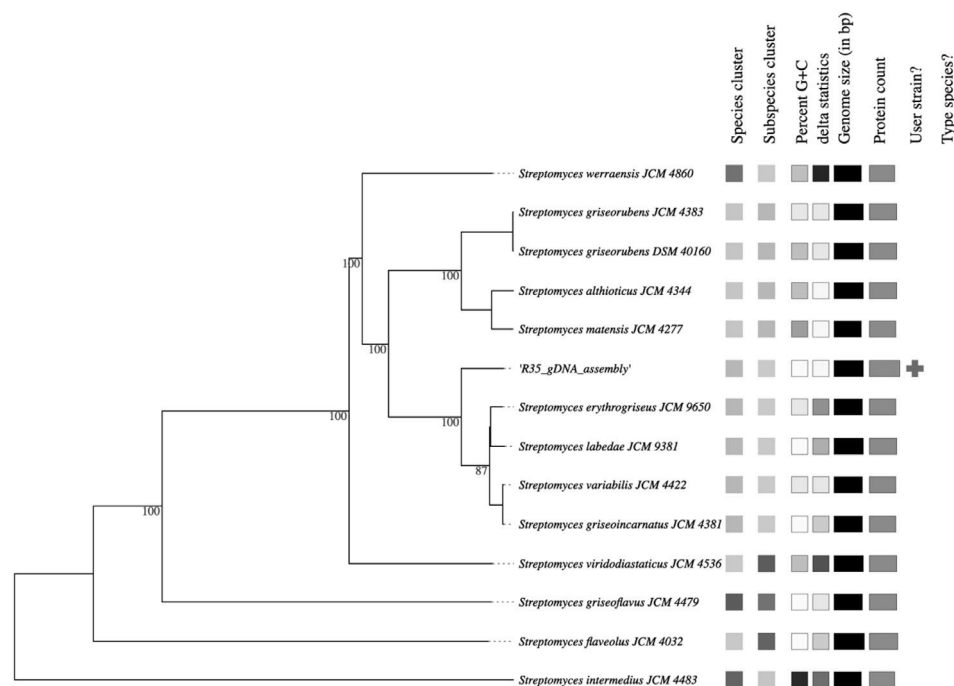


Fig. 2. Phylogram illustrating the relatedness of *Streptomyces griseoincarnatus* strain R-35 to other known *Streptomyces* strains.

35_AMR_CARD_Proksee_cgview.json). Several vancomycin and tetracycline resistance genes were observed (Fig. 3).

The presence of biosynthetic gene clusters (BGCs) was determined using antiSMASH ([file: 006_antiSMASH > index.html](http://file:006_antiSMASH>index.html)). A total of 27 BGCs were detected, predominantly polyketide synthases (PKS – type 1, type 2, and type 3), non-ribosomal peptide synthetase (NRPSs), ribosomally synthesized and post-translationally modified peptide (RiPP)-like, terpenes and lactones (Fig. 4). Sequence homology ranged from 5 to 100% similarity to known BGCs.

In addition to BGCs, carbohydrate-active enzymes (CAZymes) were determined using db-CAN3. Assignments were filtered by the predictions made by all three algorithms used. A total of 189 CAZymes were detected by all three algorithms ([file: 007_CAZymes > R-35_CAZymes_Summary.txt](http://file:007_CAZymes>R-35_CAZymes_Summary.txt)). The detected CAZymes include auxiliary activities (AA), carbohydrate-binding modules (CBM), carbohydrate esterases (CE), glycoside hydrolases (GH), glycosyltransferases (GT), and polysaccharide lyases (PL) (Fig. 5).

4. Experimental Design, Materials and Methods

All reagents were purchased from Merck-Millipore, South Africa, unless otherwise indicated. Genomic DNA was isolated from a 5-day-old liquid culture of *Streptomyces griseoincarnatus* strain R-35 cultivated in International *Streptomyces* Project medium No.2 (ISP2; g/L: 4.0 glucose; 10.0 malt extract; 4.0 yeast extract) prepared with artificial seawater (g/L: 38.2 Red Sea Salt, Red Sea™) (orbital shaking at 160rpm at 30°C). Cell mass was collected by centrifugation at 10,000rpm for 2 min. After the removal of the supernatant fluid, the cell mass was used for DNA extraction using the Quick-DNA Fecal/Soil Microbe Miniprep Kit (Zymo Research), according to the manufacturer's instructions. DNA integrity was confirmed by agarose gel electrophoresis

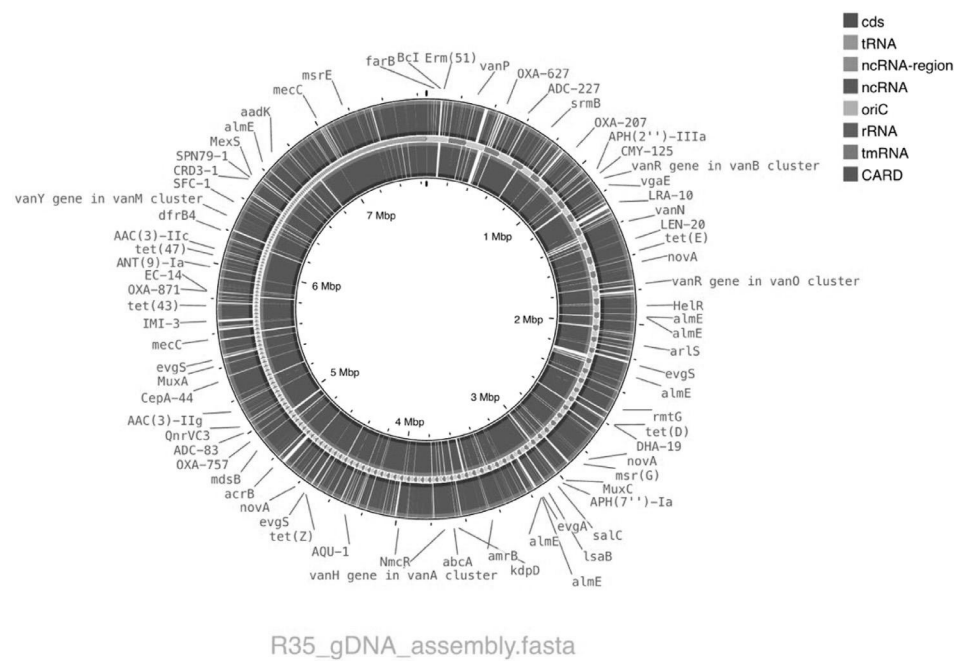


Fig. 3. Filtered circular map of antimicrobial resistance (AMR) genes detected in the draft genome of *Streptomyces griseoincarnatus* strain R-35 constructed in Proksee.

Select genomic region: Overview 2.1 3.1 5.1 6.1 18.1 22.1 24.1 28.1 30.1 36.1 43.1 45.1 47.1 61.1 64.1 100.1 101.1 110.1 118.1 122.1 173.1 236.1 260.1 271.1 278.1 279.1 387.1

Region	Type	From	To	Most similar known cluster	Similarity	Compact view
Region 2.1	T3PKS of	1	30,331	alkylresorcinol of	100%	
Region 3.1	lanthipeptide-class-I of	23,490	48,809			
Region 5.1	ectoine of	99,676	107,140	ectoine of	100%	
Region 6.1	RPP-like of	48,923	60,302	streptomycin of	75%	
Region 18.1	Ni-siderophore of	19,260	50,523	pauromycin of	9%	
Region 22.1	RPP-like of	1	7,204	informalpeptin of	28%	
Region 24.1	thiopeptide of, phenazine of, LAP of, RHE-containing of	1	43,996	oryzaniaphthopyran Aoryzaniaphthopyran Boryzaniaphthopyran Coryzanthrone Aoryzanthrone Boryzanthrone Aichlorozanthrone B of	6%	
Region 28.1	other of, NRPS of, NRPS-like of, T1PKS of	1	47,483	polyoxypyrrolin of	45%	
Region 30.1	NRPS-like of, NRPS of, PKS-like of	1	42,430	naphthylidenomycin of	64%	
Region 36.1	NRPS of, T1PKS of	3,541	38,283	antimycin of	31%	
Region 43.1	lanapeptide of	1	21,578	aborycin of	100%	
Region 45.1	ectoine of	22,243	32,641	ectoine of	100%	
Region 47.1	T3PKS of	1	35,987	spore pigment of	83%	
Region 61.1	NRPS of, lanthipeptide-class-I of	1	30,796			
Region 64.1	butyrolactone of	18,566	25,615	isoleucic acid of	17%	
Region 100.1	lanthipeptide-class-II of	7,729	23,592	SapB of	100%	
Region 101.1	terpene of	9,296	23,554	alballavenone of	100%	
Region 113.1	hydrogen-cyanide of	11,364	21,917	aborycin of	28%	
Region 119.1	RPP-like of	175	11,428			
Region 122.1	betalactone of	1	20,347			
Region 173.1	Ni-siderophore of	1	14,434	desferrioxamin Bdesferrioxamin E of	100%	
Region 236.1	terpene of	1	9,900	hopene of	53%	
Region 260.1	T1PKS of	1	8,742	primycin of	21%	
Region 271.1	NRPS of	1	8,010	incarnatapeptin Aincarnatapeptin Bidentigerumycin Fidentigerumycin G of	10%	
Region 278.1	terpene of	1	7,921	carotenoid of	49%	
Region 279.1	terpene of	1	7,744			
Region 387.1	butyrolactone of	1	3,680	griseoviridin/rimycin A of	5%	

Fig. 4. antiSMASH output of biosynthetic gene clusters detected in the draft genome assembly of *Streptomyces griseoincarnatus* strain R-35.

(1%, w/v) in 1× Tris-Acetate-EDTA buffer, pH8.3, 100V for 60 min. The gel (containing 10µg/mL ethidium bromide) was visualized using a Gel Doc XR+ (Bio-Rad) gel documentation system. The genomic DNA was submitted to the DNA Sequencing Unit at the Central Analytical Facility (CAF, Stellenbosch University, South Africa) for sequencing. The library preparation was performed in-house at CAF using the Ion 540™ Chef Kit. Sequencing was performed on the Ion Torrent GeneStudio™ S5 system using 200bp chemistry with a 100× coverage, assuming a genome size of 8-10Mbp. Raw sequence reads (.bam format) obtained from CAF were uploaded to the

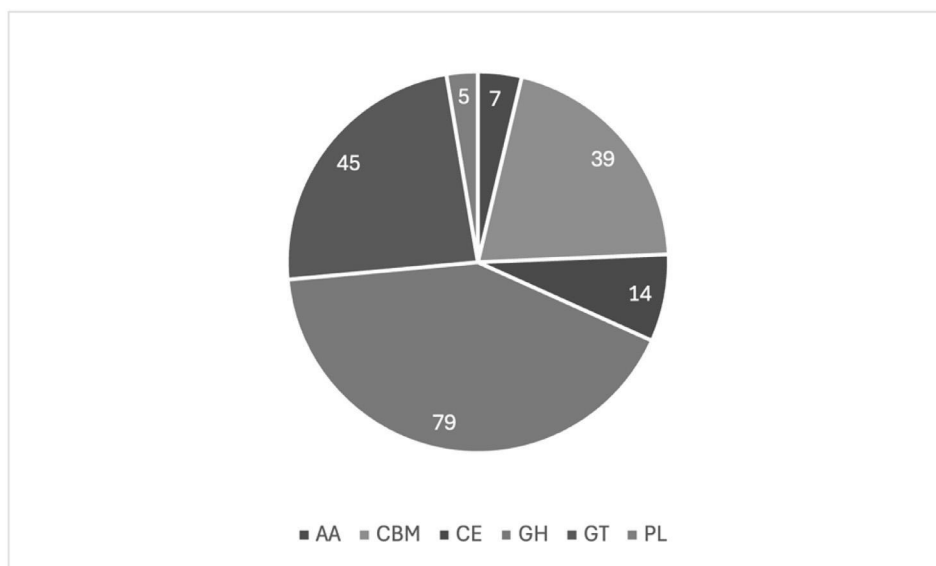


Fig. 5. Distribution of carbohydrate-active enzymes (CAZymes) detected in the draft genome of *Streptomyces griseoincarnatus* strain R-35 using dbCAN3. Data is filtered to include CAZymes only confirmed by all three algorithms. AA = auxiliary activities; CBM = carbohydrate-binding modules; CE = carbohydrate esterases; GH = glycoside hydrolases; GT = glycosyltransferases; PL = polysaccharide lyases.

Lengau cluster at the Centre for High Performance Computing (CSIR, Rosebank, South Africa). Draft genome assembly was performed using the St. Petersburg Genome Assembler (SPAdes, version 3.15.4, [8]) using an IonTorrent flag (file: 001_Assembly > R-35_gDNA_assembly.log). The resulting contigs (.fasta format) were used for downstream processing.

Default parameters were used for all bioinformatics tools unless otherwise stated. The quality of the draft genome was determined using Quast (version 5.2.0, [9]). Genome completeness and contamination were analysed using CheckM (v1.10.18) [10] on the KBase webserver (<https://kbase.us/>, [11]) with advanced parameters “Reference tree = full tree” and “Save All Plots = Save”. The Type (Strain) Genome Server (<https://tygs.dsmz.de/>, [12]) and ANI calculator (<https://www.ezbiocloud.net/tools/ani/>, [13]) were used to determine the taxonomic assignment of strain R-35. For the OrthoANI calculation, the genome sequence of the type strain, *S. griseoincarnatus* JCM 4381^T, was used for comparison with the R-35 genome. Genome annotations were performed on the Rapid Annotation using Subsystem Technology (RAST) webserver (<https://rast.nmpdr.org/>, [14]) using RASTtk, disabling the replication option, and specifying the use of the bacterial code for annotation. Additional annotations were performed with the Comprehensive Genome Analysis Tool on the BV-BRC webserver (<https://www.bv-brc.org/app/ComprehensiveGenomeAnalysis>, [15]) using the default parameters (“Assembly strategy” set to “Auto”; “Annotation recipe” set to “Bacteria / Archaea”). Additional annotations for antimicrobial resistance genes were obtained using the CARD Resistance Gene Identifier tool (version 6.0.3) on the Proksee webserver [16], enabling the options related to “loose hits”, “nudged hits” and “low sequence quality” to include shorter contigs in the draft genome assembly. The presence of biosynthetic gene clusters was determined using the antiSMASH webserver (version 7.1.0; <https://antismash.secondarymetabolites.org/>, [17]) on “relaxed” mode and enabling all additional features. Carbohydrate-active enzyme (CAZyme) annotations were obtained using the dbCAN3 webserver (sequence type was specified as “nucleotide sequence”, all tools, including “substrate prediction” were selected for the annotation) [18]. CAZymes were filtered based on the prediction by all three algorithms (HMMer, dbCAN_sub, and DIAMOND), as provided in the output table.

Limitations

The use of short-read technology (Ion Torrent) for the *de novo* assembly of the draft genome resulted in the generation of many contigs (> 500). As such, some contigs may result in suboptimal downstream processing results, particular in the case of BGC annotation.

Ethics Statement

The authors have read and followed the ethical requirements for publication in Data in Brief and hereby confirm that the current work does not involve human subjects, animal experiments, or any data collected from social media platforms.

Data Availability

Genomics dataset of marine isolate *Streptomyces griseoincarnatus* strain R-35 (Original data) (Mendeley Data).

Multi-omics data from marine sites along the Western Cape, South Africa (Original data) (NCBI).

CRediT Author Statement

Danielle Dana Mitchell: Investigation, Formal analysis, Conceptualization, Writing – review & editing; **Jo-Marie Vreulink:** Investigation, Formal analysis, Conceptualization, Writing – review & editing, Supervision; **Alaric Prins:** Investigation, Formal analysis, Conceptualization, Data curation, Writing – original draft; **Marilize Le Roes-Hill:** Investigation, Formal analysis, Conceptualization, Writing – review & editing, Supervision, Funding acquisition.

Acknowledgments

This research was supported by the National Research Foundation (NRF) of South Africa, Marine and Coastal Grant UID: 105873. The authors wish to thank the South African National Parks (SANParks) for providing a sampling permit (CRC/2019-2020/013–2019/V1). Opinions expressed, and conclusions arrived at are those of the authors and are not necessarily to be attributed to the NRF.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

References

- [1] A.C. Parte, LPSN—list of prokaryotic names with standing in nomenclature, *Nucleic Acids Res.* 42 (2014) D613–D616, doi:10.1093/nar/gkt1111.
- [2] K.F. Chater, Recent advances in understanding *Streptomyces*, *F1000Res* 5 (2016) 2795, doi:10.12688/f1000research.9534.1.
- [3] J.W.-F. Law, K.-X. Tan, S.H. Wong, N.-S.A. Mutalib, L.-H. Lee, Taxonomic and characterization methods of *Streptomyces*: a review, *Prog. Microbes. Mol. Biol.* 1 (2018), doi:10.36877/pmmmb.a0000009.
- [4] G.A. Quinn, A.M. Banat, A.M. Abdelhameed, I.M. Banat, *Streptomyces* from traditional medicine: sources of new innovations in antibiotic discovery, *J. Med. Microbiol.* 69 (2020) 1040–1048, doi:10.1099/jmm.0.001232.

- [5] F. Chassagne, G. Cabanac, G. Hubert, B. David, G. Marti, The landscape of natural product diversity and their pharmacological relevance from a focus on the dictionary of natural products®, *Phytochem. Rev.* 18 (2019) 601–622, doi:10.1007/s11101-019-09606-2.
- [6] S.S.U. Hassan, A.L. Shaikh, Marine actinobacteria as a drug treasure house, *Biomed. Pharmacother.* 87 (2017) 46–57, doi:10.1016/j.biopha.2016.12.086.
- [7] K. Alam, A. Mazumder, S. Sikdar, Y.-M. Zhao, J. Hao, C. Song, Y. Wang, R. Sarkar, S. Islam, Y. Zhang, A. Li, *Streptomyces*: the biofactory of secondary metabolites, *Front. Microbiol.* 13 (2022) 968053, doi:10.3389/fmicb.2022.968053.
- [8] A. Pribelski, D. Antipov, D. Meleshko, A. Lapidus, A. Korobeynikov, Using SPAdes De Novo assembler, *Curr. Protoc. Bioinform.* 70 (2020) e102, doi:10.1002/cpbi.102.
- [9] A. Gurevich, V. Saveliev, N. Vyahhi, G. Tesler, QUAST: quality assessment tool for genome assemblies, *Bioinformatics* 29 (2013) 1072–1075, doi:10.1093/bioinformatics/btt086.
- [10] D.H. Parks, M. Imelfort, C.T. Skennerton, P. Hugenholtz, G.W. Tyson, CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes, *Genome Res.* 25 (2015) 1043–1055, doi:10.1101/gr.186072.114.
- [11] A.P. Arkin, R.W. Cottingham, C.S. Henry, N.L. Harris, R.L. Stevens, S. Maslov, P. Dehal, D. Ware, F. Perez, S. Canon, M.W. Sneddon, M.L. Henderson, W.J. Riehl, D. Murphy-Olson, S.Y. Chan, R.T. Kamimura, S. Kumari, M.M. Drake, T.S. Brettin, E.M. Glass, D. Chivian, D. Gunter, D.J. Weston, B.H. Allen, J. Baumohl, A.A. Best, B. Bowen, S.E. Brenner, C.C. Bun, J.-M. Chandonia, J.-M. Chia, R. Colasanti, N. Conrad, J.J. Davis, B.H. Davison, M. DeJongh, S. Devoid, E. Dietrich, I. Dubchak, J.N. Edirisinghe, G. Fang, J.P. Faria, P.M. Frybarger, W. Gerlach, M. Gerstein, A. Greiner, J. Gurtowski, H.L. Haun, F. He, R. Jain, M.P. Joachimiak, K.P. Keegan, S. Kondo, V. Kumar, M.L. Land, F. Meyer, M. Mills, P.S. Novichkov, T. Oh, G.J. Olsen, R. Olson, B. Parrello, S. Pasternak, E. Pearson, S.S. Poon, G.A. Price, S. Ramakrishnan, P. Ranjan, P.C. Ronald, M.C. Schatz, S.M.D. Seaver, M. Shukla, R.A. Sutormin, M.H. Syed, J. Thomason, N.L. Tintle, D. Wang, F. Xia, H. Yoo, S. Yoo, D. Yu, KBase: the United States department of energy systems biology knowledge-base, *Nat. Biotechnol.* 36 (2018) 566–569, doi:10.1038/nbt.4163.
- [12] J.P. Meier-Kolthoff, M. Göker, TYGS is an automated high-throughput platform for state-of-the-art genome-based taxonomy, *Nat. Commun.* 10 (2019) 2182, doi:10.1038/s41467-019-10210-3.
- [13] S.H. Yoon, S.M. Ha, J.M. Lim, S.J. Kwon, J. Chun, A large-scale evaluation of algorithms to calculate average nucleotide identity, *Antonie van Leeuwenhoek* 110 (2017) 1281–1286, doi:10.1007/s10482-017-0844-4.
- [14] 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 Genom.* 9 (2008) 75, doi:10.1186/1471-2164-9-75.
- [15] R.D. Olson, R. Assaf, T. Brettin, N. Conrad, C. Cucinell, J.J. Davis, D.M. Dempsey, A. Dickerman, E.M. Dietrich, R.W. Kenyon, M. Kucuglu, E.J. Lefkowitz, J. Lu, D. Machi, C. Macken, C. Mao, A. Niewiadomska, M. Nguyen, G.J. Olsen, J.C. Overbeek, B. Parrello, V. Parrello, J.S. Porter, G.D. Pusch, M. Shukla, I. Singh, L. Stewart, G. Tan, C. Thomas, M. VanOeffelen, V. Vonstein, Z.S. Wallace, A.S. Warren, A.R. Wattam, F. Xia, H. Yoo, Y. Zhang, C.M. Zmasek, R.H. Scheuermann, R.L. Stevens, Introducing the bacterial and viral bioinformatics resource centre (BV-BRC): a resource combining PATRIC, IRD and ViPR, *Nucl. Acids Res.* 51 (2023) D678–D689, doi:10.1093/nar/gkac1003.
- [16] J.R. Grant, E. Enns, E. Marinier, A. Mandal, E.K. Herman, C. Chen, M. Graham, G. Van Domselaar, P. Stothard, Proksee: in-depth characterization and visualization of bacterial genomes, *Nucl. Acids Res.* 51 (2023) W484–W492, doi:10.1093/nar/gkad326.
- [17] K. Blin, S. Shaw, H.E. Augustijn, Z.L. Reitz, F. Biermann, M. Alanjary, A. Fetter, B.R. Terlouw, W.W. Metcalf, E.J.N. Helfrich, G.P. van Wezel, M.H. Medema, T. Weber, antiSMASH 7.0: new and improved predictions for detection, regulation, chemical structures and visualisation, *Nucl. Acids Res.* 51 (2023) W46–W50, doi:10.1093/nar/gkad344.
- [18] J. Zheng, Q. Ge, Y. Yan, X. Zhang, L. Huang, Y. Yin, dbCAN3: automated carbohydrate-active enzyme and substrate annotation, *Nucl. Acids Res.* 51 (2023) W115–W121, doi:10.1093/nar/gkad328.