ACCESS MICROBIOLOGY
an open research platform

SHORT COMMUNICATION
Borba *et al.*, *Access Microbiology* 2023;5:000466.v3
DOI 10.1099/acmi.0.000466.v3

MICROBIOLOGY SOCIETY
OPEN DATA OPEN ACCESS

Whole-genome sequencing-based characterization of *Streptomyces* sp. 6(4): focus on natural product

Marcela Proença Borba[1,*], João Paulo Witusk[1], Débora Marchesan Cunha[1], Daiana de Lima-Morales[2,3], Andreza Francisco Martins[1] and Sueli Van Der Sand[1]

Abstract

We have sequenced the whole genome of *Streptomyces* sp. 6(4) isolated from tomato roots that presents antifungal activity against phytopathogenic fungi, mainly *Bipolaris sorokiniana*. The genome has almost 7 Mb and 3368 hypothetical proteins that were analysed and characterized in Uniprot with the emphasis on biological compounds. Multilocus sequence typing (MLST) analyses were performed in an effort to characterize and identify this isolate, resulting in a new sequence type (ST), classified as ST64. Phenetic and phylogenetic trees were constructed to investigate *Streptomyces* sp. 6(4) evolution and sequence similarity, and the isolate is a strain closer to *Streptomyces prasinus* and *Streptomyces viridosporus*. It is known that the genus *Streptomyces* possess huge metabolic capacity with the presence of cryptic genes. These genes are usually present in clusters, which are responsible for the production of diverse natural products, mainly antibiotics. In addition, 6(4) showed 11 biosynthetic gene clusters through antiSMASH, including 3 polyketide synthase (PKS) and non-ribosomal peptide synthase (NRPS) type clusters.

DATA SUMMARY

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession VIFW00000000. Due to the large number of nucleotide sequences, the database accession numbers are found in tables throughout the paper and in the supplementary material.

INTRODUCTION

The genus *Streptomyces* comprises Gram-positive bacteria that are commonly found in the soil and produce a great number of secondary metabolites, many of which have distinct biological activities. These natural products are well known for their antimicrobial activities, and *Streptomyces* compounds comprise 70% of the antibiotics used in medicine [1, 2]. The discovery of new antibiotics from bacterial sources has been limited to the reisolation of the same molecules by traditional methods [3]. Fortunately, the genome sequencing of Actinomycetes lineages since Bentley *et al.* [4] has demonstrated biosynthetic genes that can be silenced in certain situations but are capable of producing bioactive metabolites [5]. This metabolic richness defined a new basis for natural product research, based on genome sequencing and mining [6, 7], demonstrating that this genus is still a promising source of new compounds.

The genus *Streptomyces* has a problematic history with respect to species identification, where strains with different phenotypes, morphology and biochemistry have identical 16S rRNA sequences [8]. DNA hybridization and DNA fingerprinting were used in phylogenetics studies, but no one has found suitable resolution and reproducibility for identification worldwide [9]. Whole-genome sequencing seems to be the best available technique to identify species in the genus *Streptomyces*. Here, we sequenced the whole genome of *Streptomyces* sp. 6(4), an isolate from our research group that produces antifungal compounds and has been studied for a long time without proper species definition.

Received 27 June 2022; Accepted 23 September 2022; Published 03 March 2023
Author affiliations: [1]Programa de Pós-graduação em Microbiologia Agrícola e do Ambiente, Instituto de Ciências Básicas da Saúde, Universidade Federal do Rio Grande do Sul, Porto Alegre, Brazil; [2]Núcleo de Bioinformática do Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil; [3]Laboratório de Pesquisa em Resistência Bacteriana (LABRESIS), Hospital de Clínicas de Porto Alegre, Porto Alegre, Rio Grande do Sul, Brazil.
***Correspondence: Marcela Proença Borba,** ceh.proenca@gmail.com
Keywords: Actinomycetes; biosynthetic gene cluster; genome mining; phytopathogenic fungi; secondary metabolites.
Abbreviations: atpD, ATP synthase subunit beta; gyrB, DNA gyrase subunit B; recA, Recombinase A; rpoB, RNA polymerase subunit beta; trpB, Tryptophan synthase beta chain.
Three supplementary figures and three supplementary tables are available with the online version of this article.

Table 1. Clusters and proteins identified in *Streptomyces* 6(4) analysis in antiSMASH

Region	Type	Proteins
4.1	Melanin	Tyrosinase
		Cytochrome P450
29.1	Arylpolyene	Hypothetical protein
		3-oxoacyl-ACP synthase
		o-succunylbenzoate--CoA ligase
100.1	Terpene	Sporulation protein
		Hypothetical protein
		Lycopene cyclase
140.1	RiPP-like	Endonuclease
165.1	PKS-like	Endonuclease
		Type I polyketide synthase
		Acyl-carrier-protein S-malonyltransferase
186.1	Terpene	Phytoene synthase
		2Fe-2S ferredoxin
207.1	RiPP-like	Hypothetical protein
		Endonuclease
243.1	NRPS-like	Serine hydroxymethyltransferase
		Non-ribosomal peptide synthetase
268.1	T1PKS	Type I polyketide synthase
276.1	NRPS	Non-ribosomal peptide synthetase
327.1	NRPS	Non-ribosomal peptide synthetase

METHODS

Data collection

Streptomyces sp. 6(4) was previously isolated from tomato plant root (*Lycopersicon esculentum* from 29°55′8″S, 51°10′41″W) and has been partially identified morphologcally and biochemically [10]. This isolate produces compounds against Gram-negative and Gram-positive bacteria, besides yeasts and especially filamentous fungi [11]. The DNA was obtained by phenol–chloroform extraction and the strain was stocked in agar discs in 20% glycerol at −20 °C.

Whole-genome sequencing (WGS)

The MiSeq platform (Illumina, San Diego, CA, USA) was used for WGS of the streptomycetes isolate. The paired-end library was performed with the Nextera XT DNA Library Prep kit (Illumina). The run was performed with the MiSeq Reagent v2 kit (500 cycles) with a calculated coverage of 100×. *De novo* assembly of the genome was performed using SPAdes (v3.6.2; http://cab.spbu.ru/software/spades/). The file generated by the assembly was annotated on the RAST platform (http://rast.nmpdr.org) and Prokka (contigs ≥200 bp; ≥10× coverage) [12]. The *Streptomyces* sp. 6(4) genome was visualized in Geneious Prime 9.0.5. The hypothetical protein annotation was performed after BLASTX (National Center for Biotechnology Information, NCBI) searches and UniProt BLAST analysis [13]. antiSMASH 4.0 software [14] was used to predict biosynthetic gene clusters for secondary metabolites.

The six alleles, 16S rRNA, *trpB*, *recA*, *rpoB*, *gyrB* and *atpD*, were available *in silico* and the sequence type (ST) assignments were defined using the *Streptomyces* database at PubMLST [15].

Phylogenetic and phenetic trees

Multilocus sequence typing (MLST)-based phylogenetic and phenetic trees were developed by aligning concatenated sequences of six conserved genes (16S rRNA, *rpoB*, *trpB*, *recA*, *gyrB* and *atpD*). Gene alignment was performed in Geneious Prime 9.0.5 using ClustalW. Two different algorithms and evolutionary models were used, neighbour joining with the Kimura two-parameter

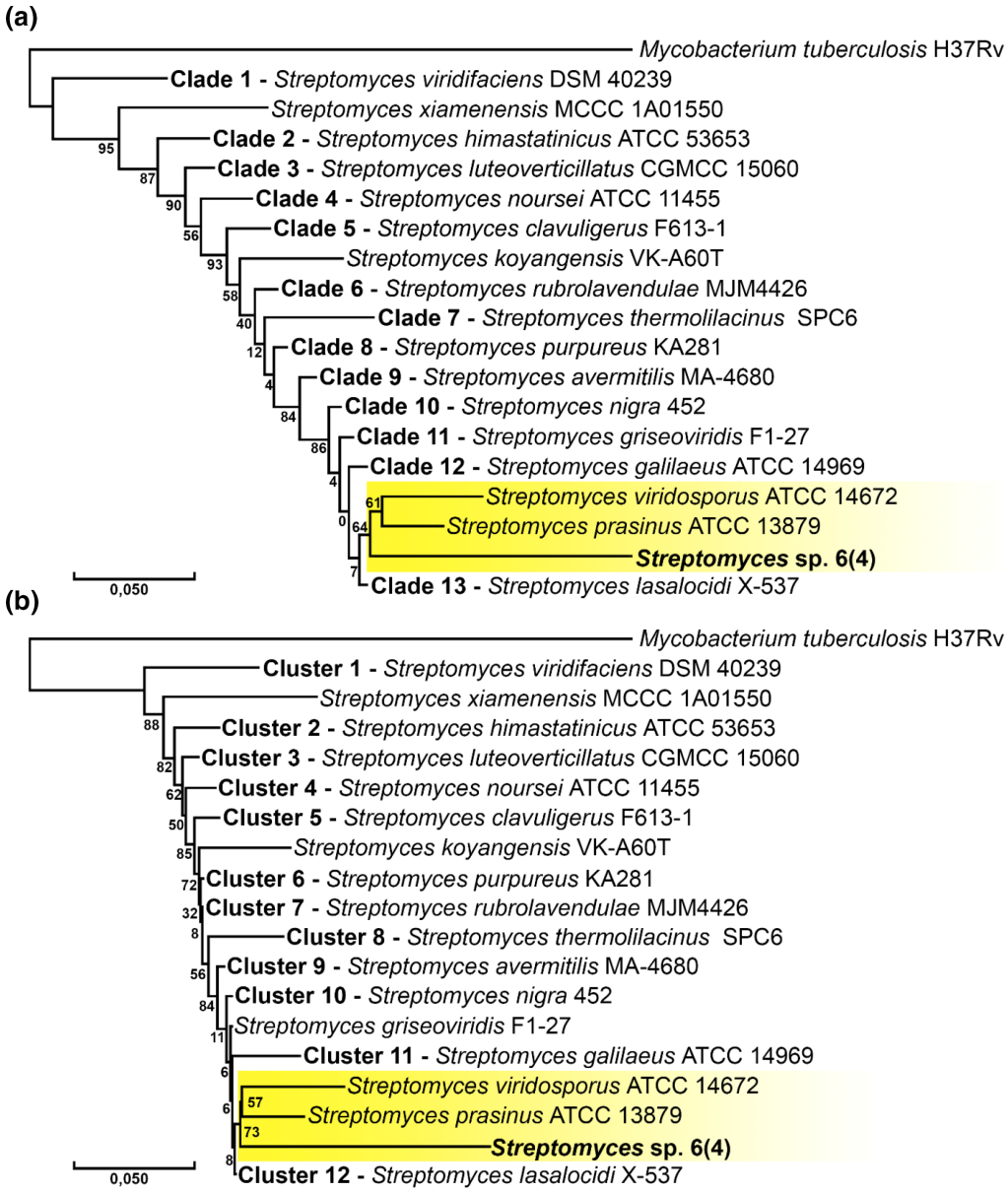

Fig. 1. Maximum-likelihood phylogenetic tree with the general time-reversible model (a) and neighbour-joining phenetic tree with the Kimura two-parameter model (b) for *Streptomyces* sp. 6(4) with 102 *Streptomyces* representative genome strains coalesced into clusters for a clearer view (see Figs S2 and S3). *M. tuberculosis* H37Rv is used as an external group (Table S3).

model and maximum likelihood with the general time-reversible model, both with 500 bootstrap replicates with 102 no-trimmed sequences from *Streptomyces* representative genome strains from the NCBI using *Mycobacterium tuberculosis* H37Rv as an external group in MEGAX software [16]. For brevity, non-target clades and groups were coalesced in the figures to highlight the position of 6(4) in both analyses; complete trees are available in Figs S2 and S3 (available in the online version of this article).

A new *Streptomyces* genome related to *S. prasinus* and *S. viridosporus* was deposited in GenBank and the *Streptomyces* sp. 6(4) is able to generate natural products according to its biosynthetic gene clusters.

RESULTS AND DISCUSSION

This Whole Genome Shotgun project has been deposited at DDBJ/ENA/GenBank under the accession VIFW00000000. The complete linear genome was 6 868 166 bp with 72.7% GC content (2477 contigs; N50=7987). Six thousand and thirty-eight protein-coding genes and 76 RNA genes were detected. All 3368 hypothetical proteins were analysed and only 1132 hypothetical

proteins had their function identified in Uniprot. Protein analysis in BLASTX showed that 60 hypothetical proteins are involved in the metabolism process and 72 are involved in biosynthetic metabolism according to Uniprot (Table S1).

With this in view, we identified some proteins in 6(4) that are involved in biosynthetic metabolism and could be investigated in further studies. One of these is a 4,5-DOPA dioxygenase estradiol that is part of the construction process of various biological agents, such as anthramycin, porothramycin, sibiromycin, tomaymycin, hormaomycins and lincomycin A [17]. An antibiotic biosynthesis monooxygenase protein was recognized in 6(4) and seems to be part of some already known pathways of antifungal and antibiotic synthesis, such as nystatin in *Streptomyces noursei* [18] and actinorhodin in *Streptomyces coelicolor* [19], respectively. The thiosterase is another protein that has been identified and it is part of polyketide synthase (PKS) construction [20, 21]. The PKSs and non-ribosomal peptide synthases (NRPSs) are enzymatic complexes that are responsible for the formation of polyketides and non-ribosomal peptides that form two large and important groups of natural products with different chemical structures responsible for a wide variety of biologically active compounds [22]. Another important enzyme involved in the construction of PKSs and NRPSs is the aminotransferase class I/II-fold pyridoxal phosphate-dependent enzyme [23], present in the 6(4) isolate. We also identified a thiazolylpeptide-type bacteriocin. Thiazolyl peptides are antibiotics originated by translation modifications from ribosomal natural products and are found in streptomycetes [24], and thiazole moiety is used in all penicillin derivatives such as thiabendazole, a known fungicide [25]. Terpenoids are a diverse group of natural products with antifungal properties [26], commonly found in plants, and need two initial primary metabolites to be produced, isopentenyl diphosphate (IPP) and dimethylallyl diphosphate (DMAPP). Both carbon units are originated by distinct pathways. The 3-hydroxy-3-methylglutaryl-ACP synthase, observed in 6(4), is present in the genus *Streptomyces* and is responsible for terpenoid biosynthesis by the mevalonate pathway [27]. Squalene synthase HpnD is also a protein required by the mevalonate pathway to produce squalene, a kind of terpenoid from the isoprenoid group [28]. Another independent pathway for terpenoids involves 1-deoxy-D-xylulose-5-phosphate synthase, which condenses pyruvate and glyceraldehyde 3-phosphate [29, 30]. Both pathways seem to be present in 6(4). Anthranilate phosphoribosyltransferase is involved in the biosynthesis of a nonribosomal lipopeptide called a calcium-dependent antibiotic (CDA) in, at least, *S. coelicolor* [31], as with nucleotide sugar dehydrogenase that is the precursor of congocidine biosynthesis [32]. Both enzymes were found in this 6(4)isolate. Other proteins were identified in this study as biosynthesis proteins. Many of them were part of carbohydrate metabolism, acid folic metabolism, amino acids and vitamins. In these cases, nothing was uncovered about biological compounds, but this does not mean that these enzymes do not participate in natural product production; we simply do not know yet, and it is very common for these metabolism pathways to generate active compounds by some means. Further, some proteins described above seems to be part of diverse known antifungal products, which can corroborate the antifungal action of 6(4) against *B. sorokiniana*.

The first streptomycete that has had its whole genome sequenced – because of important bioactive compounds derived from secondary metabolites [4] – belongs to *S. coelicolor*. The authors have sequenced the genome of *S. coelicolor* A3(2) and discovered just by analysing the genome that, besides the 5 compounds already known, the isolate could produce approximately 20 different compounds. Bentley's study showed that this group of bacteria is still a source for the discovery of new metabolites with biological activities. This and other studies [33, 34] showed the existence of cryptic genes that are silent in certain conditions and are the missing gap in the search for new natural products. With the whole genome sequence, genome mining is used to find the biosynthetic gene clusters (BGCs) in *Streptomyces*, known for the large number of these clusters in their genomes. These BGCs are categorized into 34 major classes [35]. In antiSMASH we identified 11 gene clusters encoding pathways for secondary metabolism. All proteins have been searched in Geneious software and BLASTP to confirm the mining (Table 1).

Two clusters are terpene types. The terpenoids are famous natural products and can be produced by the mevalonate pathway, as seen above. These terpenoids are widely produced by *Streptomyces* [36] and this strain can probably produce them too. The BGC in the 186.1 region, found in 6(4), showed sequence similarity with terpene BGCs in *Streptomyces griseus* subsp. *griseus* NBRC 13350 (BGC000664 [37]), *Streptomyces griseus* (BGC0000649 [38]), *Streptomyces collinus* (BGC0001227 [39]), *Streptomyces scabiei* (BGC0001456 [40]) and *Streptomyces avermitilis* (BGC0000633 [41]) (Fig. S1). Aryl polyene in region 29.1 is a pigment polyketide originated by the PKS system [42] and at least another three clusters encode PKS and NRPS genes, an important discovery indicating that 6(4) has distinct antimicrobial activity, e.g. antifungal, acting against a variety of micro-organisms, as shown in later studies [11].

AntiSMASH analysis confirmed that this strain has different gene clusters and can produce different biological compounds. Although it is very common to find hopene in the *Streptomyces* genome because of aerial growth of the mycelia, in 6(4) it was not possible to detect it. This is not a problem due to the fact that not all streptomycetes necessarily seem to produce this compound [43]. However, we have not been able to identify geosmin and this could be due to some mishap in the sequencing process.

The genome was analysed in PubMLST and classified as ST64, a new ST deposited in the website as id 271.

A phylogenetic tree was constructed to infer evolutionary relationships with the maximum-likelihood algorithm (Fig. 1a) and a phenetic tree was constructed to observe sequence similarities, using neighbour joining (Fig. 1b). Both trees aim to characterize the 6(4) and are commonly used to deduce species identification in microbiology. In this study, the *Streptomyces* sp. 6(4) is related

to *S. prasinus* ATCC 13879 and *S. viridosporus* ATCC 14672, forming a sister group to these taxa in both attempted trees (Fig. 1), showing high similarity and a close relationship with these two species.

Morphological description and MLSA analysis of *S. prasinus* were presented by Labeda *et al.* [44] and it was found that *S. prasinus* shared morphological traits with *S. viridosporus* according to Goodfellow *et al.* [45]. Although the 16S rRNA showed high similarity with *Streptomyces griseoincarnatus* when analysed in EzBioCloud [46], in the NCBI database it showed similarity with *Streptomyces griseoflavus*. The MLST genes were investigated in the NCBI database, but lower percentage rates were achieved. Some species (*Streptomyces cadmiisoli, S. collinus, Streptomyces hygroscopicus, S. incarnatus* and *Streptomyces olivaceus*) appear in all housekeeping genes. *S. coelicolor, Streptomyces albogriseolus* and *Streptomyces ambofaciens* are species that also appear in all housekeeping genes and in the 16S rRNA, but with different similarity percentages (Table S2). Even with the complete genome sequenced, we cannot define the species that fits this isolate genomically, as usual with members of the genus *Streptomyces*, due to the large number of species and poor variation in 16S rRNA [47]. The publication of new species belonging to the genus requires several methods in order to reach a final conclusion [48], since the methods separately are not sufficient even for the exact identification of already known species. This is due to the large number of species attributed to the genus – approximately 3000 different species – especially in patent records [49]. Today, searching for *Streptomyces* at www.bacterio.net, we arrived at 696 species in validated publications and corrected names without synonyms, which is still a large number of species belonging to the same genus, making the taxonomy of streptomycetes more complicated than in other bacterial genera.

This paper also aims to show the variety of metabolic possibilities in the genome of one *Streptomyces* isolate but we cannot affirm that any biosynthetic gene cluster present in 6(4) is activated just because the isolate has activity against phytopathogenic fungi.

Funding information
This study was funded by Coordination for the Improvement of Higher Education Personnel (CAPES) (88882.439723/2019–01), National Council for Scientific and Technological Development (CNPq) and National Institute of Antimicrobial Resistance Research – INPRA (MCTI/CNPq/CAPES/FAPs no. 16/2014).

Acknowledgements
The authors would like to thank the development agencies for financial support and Voltaire Dutra Paes Neto for figure artwork.

Author contribution
M. P. B.: conceptualization, methodology, investigation, data curation, writing – original draft preparation and visualization. J. P. D. W., D. M. C. and D. d. L. M.: investigation and data curation. A. F. M.: conceptualization methodology, data curation, resources, writing – review and editing and funding. S. V. Der S.: conceptualization, resources, writing – review and editing, supervision and funding.

Conflicts of interest
The authors declare that there are no conflicts of interest.

References

1. Bérdy J. Thoughts and facts about antibiotics: where we are now and where we are heading. *J Antibiot* 2012;65:385–395.
2. Hutchings MI, Truman AW, Wilkinson B. Antibiotics: past, present and future. *Curr Opin Microbiol* 2019;51:72–80.
3. Aigle B, Lautru S, Spiteller D, Dickschat JS, Challis GL, *et al*. Genome mining of *Streptomyces ambofaciens*. *J Ind Microbiol Biotechnol* 2014;41:251–263.
4. Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, *et al*. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 2002;417:141–147.
5. Nett M, Ikeda H, Moore BS. Genomic basis for natural product biosynthetic diversity in the actinomycetes. *Nat Prod Rep* 2009;26:1362–1384.
6. Corre C, Challis GL. New natural product biosynthetic chemistry discovered by genome mining. *Nat Prod Rep* 2009;26:977–986.
7. Zerikly M, Challis GL. Strategies for the discovery of new natural products by genome mining. *Chembiochem* 2009;10:625–633.
8. Antony-Babu S, Stien D, Eparvier V, Parrot D, Tomasi S, *et al*. Multiple *Streptomyces* species with distinct secondary metabolomes have identical 16S rRNA gene sequences. *Sci Rep* 2017;7:11089.
9. Guo Y, Zheng W, Rong X, Huang Y. A multilocus phylogeny of the *Streptomyces griseus* 16S rRNA gene clade: use of multilocus sequence analysis for streptomycete systematics. *Int J Syst Evol Microbiol* 2008;58:149–159.
10. de Oliveira MF, da Silva MG, Van Der Sand ST. Anti-phytopathogen potential of endophytic actinobacteria isolated from tomato plants (*Lycopersicon esculentum*) in southern Brazil, and characterization of *Streptomyces* sp. R18(6), a potential biocontrol agent. *Res Microbiol* 2010;161:565–572.
11. Monteiro P, Borba MP, Van Der Sand ST. Evaluation of the antifungal activity of *Streptomyces* sp. on *Bipolaris sorokiniana* and the growth promotion of wheat plants. *JAS* 2017;9:229.
12. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
13. UniProt Consortium. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res* 2021;49:D480–D489.
14. Blin K, Wolf T, Chevrette MG, Lu X, Schwalen CJ, *et al*. antiSMASH 4.0-improvements in chemistry prediction and gene cluster boundary identification. *Nucleic Acids Res* 2017;45:W36–W41.
15. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome Open Res* 2018;3:124.
16. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms. *Mol Biol Evol* 2018;35:1547–1549.
17. Zhong G, Zhao Q, Zhang Q, Liu W. 4-alkyl-L-(Dehydro)proline biosynthesis in actinobacteria involves N-terminal nucleophile-hydrolase activity of γ-glutamyltranspeptidase homolog for C-C bond cleavage. *Nat Commun* 2017;8:16109.
18. Volokhan O, Sletta H, Ellingsen TE, Zotchev SB. Characterization of the P450 monooxygenase NysL, responsible for C-10 hydroxylation during biosynthesis of the polyene macrolide antibiotic nystatin in *Streptomyces noursei*. *Appl Environ Microbiol* 2006;72:2514–2519.

19. Kendrew SG, Hopwood DA, Marsh EN. Identification of a monooxygenase from *Streptomyces coelicolor* A3(2) involved in biosynthesis of actinorhodin: purification and characterization of the recombinant enzyme. *J Bacteriol* 1997;179:4305–4310.
20. Kotowska M, Pawlik K, Smulczyk-Krawczyszyn A, Bartosz-Bechowski H, Kuczek K. Type II thioesterase scot, associated with *Streptomyces coelicolor* A3(2) modular polyketide synthase cpk, hydrolyzes acyl residues and has a preference for propionate. *Appl Environ Microbiol* 2009;75:887–896.
21. Whicher JR, Florova G, Sydor PK, Singh R, Alhamadsheh M, *et al*. Structure and function of the RedJ protein, a thioesterase from the prodiginine biosynthetic pathway in *Streptomyces coelicolor*. *J Biol Chem* 2011;286:22558–22569.
22. Du L, Lou L. PKS and NRPS release mechanisms. *Nat Prod Rep* 2010;27:255–278.
23. Milano T, Paiardini A, Grgurina I, Pascarella S. Type I pyridoxal 5'-phosphate dependent enzymatic domains embedded within multimodular nonribosomal peptide synthetase and polyketide synthase assembly lines. *BMC Struct Biol* 2013;13:26.
24. Young TS, Walsh CT. Identification of the thiazolyl peptide GE37468 gene cluster from *Streptomyces* ATCC 55365 and heterologous expression in *Streptomyces lividans*. *Proc Natl Acad Sci* 2011;108:13053–13058.
25. Eissa SI, Farrag AM, Abbas SY, El Shehry MF, Ragab A, *et al*. Novel structural hybrids of quinoline and thiazole moieties: Synthesis and evaluation of antibacterial and antifungal activities with molecular modeling studies. *Bioorg Chem* 2021;110:104803.
26. Zhang K, Jiang Y, Zhao H, Köllner TG, Chen S, *et al*. Diverse terpenoids and their associated antifungal properties from roots of different cultivars of *Chrysanthemum Morifolium* ramat. *Molecules* 2020;25:2083.
27. Kuzuyama T. Biosynthetic studies on terpenoids produced by *Streptomyces*. *J Antibiot* 2017;70:811–818.
28. Ghimire GP, Thuan NH, Koirala N, Sohng JK. Advances in biochemistry and microbial production of squalene and its derivatives. *J Microbiol Biotechnol* 2016;26:441–451.
29. Poralla K, Muth G, Härtner T. Hopanoids are formed during transition from substrate to aerial hyphae in *Streptomyces coelicolor* A3(2). *FEMS Microbiol Lett* 2000;189:93–95.
30. Kaneda K, Kuzuyama T, Takagi M, Hayakawa Y, Seto H. An unusual isopentenyl diphosphate isomerase found in the mevalonate pathway gene cluster from *Streptomyces* sp. strain CL190. *Proc Natl Acad Sci* 2001;98:932–937.
31. Hojati Z, Milne C, Harvey B, Gordon L, Borg M, *et al*. Structure, biosynthetic origin, and engineered biosynthesis of calcium-dependent antibiotics from *Streptomyces coelicolor*. *Chem Biol* 2002;9:1175–1187.
32. Juguet M, Lautru S, Francou F-X, Nezbedová S, Leblond P, *et al*. An iterative nonribosomal peptide synthetase assembles the pyrrole-amide antibiotic congocidine in *Streptomyces ambofaciens*. *Chem Biol* 2009;16:421–431.
33. Medema MH, Trefzer A, Kovalchuk A, van den Berg M, Müller U, *et al*. The sequence of a 1.8-mb bacterial linear plasmid reveals a rich evolutionary reservoir of secondary metabolic pathways. *Genome Biol Evol* 2010;2:212–224.
34. Becerril A, Álvarez S, Braña AF, Rico S, Díaz M, *et al*. Uncovering production of specialized metabolites by *Streptomyces argillaceus*: activation of cryptic biosynthesis gene clusters using nutritional and genetic approaches. *PLoS One* 2018;13:e0198145.
35. Belknap KC, Park CJ, Barth BM, Andam CP. Genome mining of biosynthetic and chemotherapeutic gene clusters in *Streptomyces* bacteria. *Sci Rep* 2020;10:2003.
36. Martín-Sánchez L, Singh KS, Avalos M, van Wezel GP, Dickschat JS, *et al*. Phylogenomic analyses and distribution of terpene synthases among *Streptomyces*. *Beilstein J Org Chem* 2019;15:1181–1193.
37. Krügel H, Krubasik P, Weber K, Saluz HP, Sandmann G. Functional analysis of genes from *Streptomyces griseus* involved in the synthesis of isorenieratene, a carotenoid with aromatic end groups, revealed a novel type of carotenoid desaturase. *Biochim Biophys Acta - Mol Cell Biol Lipids* 1999;1439:57–64.
38. Lee HS, Ohnishi Y, Horinouchi S. A sigmab-like factor responsible for carotenoid biosynthesis in *Streptomyces griseus*. *J Mol Microbiol Biotechnol* 2001;3:95–101.
39. Iftime D, Kulik A, Härtner T, Rohrer S, Niedermeyer THJ, *et al*. Identification and activation of novel biosynthetic gene clusters by genome mining in the kirromycin producer *Streptomyces collinus* Tü 365. *J Ind Microbiol Biotechnol* 2016;43:277–291.
40. Liu J, Nothias L-F, Dorrestein PC, Tahlan K, Bignell DRD. Genomic and metabolomic analysis of the potato common scab pathogen *Streptomyces scabiei*. *ACS Omega* 2021;6:11474–11487.
41. Omura S, Ikeda H, Ishikawa J, Hanamoto A, Takahashi C, *et al*. Genome sequence of an industrial microorganism *Streptomyces avermitilis*: deducing the ability of producing secondary metabolites. *Proc Natl Acad Sci* 2001;98:12215–12220.
42. Grammbitter GLC, Schmalhofer M, Karimi K, Shi Y-M, Schöner TA, *et al*. An uncommon type II PKS catalyzes biosynthesis of aryl polyene pigments. *J Am Chem Soc* 2019;141:16615–16623.
43. Seipke RF, Loria R. Hopanoids are not essential for growth of *Streptomyces scabies* 87-22. *J Bacteriol* 2009;191:5216–5223.
44. Labeda DP, Rong X, Huang Y, Doroghazi JR, Ju K-S, *et al*. Taxonomic evaluation of species in the *Streptomyces hirsutus* clade using multi-locus sequence analysis and proposals to reclassify several species in this clade. *Int J Syst Evol Microbiol* 2016;66:2444–2450.
45. Goodfellow M, Busarakam K, Idris H, Labeda DP, Nouioui I, *et al*. *Streptomyces asenjonii* sp. nov., isolated from hyper-arid Atacama Desert soils and emended description of *Streptomyces viridosporus* Pridham et al. 1958. *Antonie van Leeuwenhoek* 2017;110:1133–1148.
46. Yoon SH, Ha SM, Kwon S, Lim J, Kim Y, *et al*. Introducing EzBioCloud: a taxonomically united database of 16S rRNA gene sequences and whole-genome assemblies. *Int J Syst Evol Microbiol* 2017;67:1613–1617.
47. Labeda DP, Dunlap CA, Rong X, Huang Y, Doroghazi JR, *et al*. Phylogenetic relationships in the family *Streptomycetaceae* using multi-locus sequence analysis. *Antonie van Leeuwenhoek* 2017;110:563–583.
48. Schwitalla JW, Benndorf R, Martin K, Vollmers J, Kaster A-K, *et al*. *Streptomyces smaragdinus* sp. nov., isolated from the gut of the fungus growing-termite *Macrotermes natalensis Int J Syst Evol Microbiol* 2020;70:5806–5811.
49. Law J-F, Tan K-X, Wong SH, Ab Mutalib N-S, Lee L-H. Taxonomic and characterization methods of *Streptomyces*: a review. *Prog Micobes Mol Biol* 2018;1:a0000009.

Peer review history

VERSION 2

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000466.v2.3

John Munnoch; University of Strathclyde, SIPBS, UNITED KINGDOM, Glasgow

Date report received: 23 September 2022
Recommendation: Accept

Comments: The manuscript is well written and contributes to the literature. Thank you for addressing all reviewers comments satisfactorily and in a timely manner, this was very much appreciated.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000466.v2.1

Rebecca E McHugh; University of Glasgow Institute of Infection Immunity and Inflammation, Bacteriology, 50 Doon Crescent, Bearsden, Glasgow, UNITED KINGDOM
https://orcid.org/0000-0003-4664-6125

Date report received: 21 September 2022
Recommendation: Accept

Comments: 1. Methodological rigour, reproducibility and availability of underlying data Good, much improved. 2. Presentation of results Good. 3. How the style and organization of the paper communicates and represents key findings Good. 4. Literature analysis or discussion Good- discussion and reference to literature has been improved.

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000466.v2.2

Rebecca Devine; John Innes Centre, UNITED KINGDOM
https://orcid.org/0000-0003-0008-2184

Date report received: 19 September 2022
Recommendation: Accept

Comments: The manuscript presented describes the genome sequence of Streptomyces sp. 6(4), a strain previously isolated from tomato root that shows antifungal activity against the plant-pathogen, Bipolaris sorokiniana. The MiSeq Illumina platform was used to sequence the genomic DNA and the genes present have been annotated using several methods. Phylogenetic analysis has been conducted to identify the species of the isolate and AntiSMASH analysis to predict biosynthetic genes and gene clusters present. The authors have made some preliminary suggestions as to which of these genes/BGCs may be responsible for the previously observed bioactivity. The manuscript is generally well presented. The workflow is easy to follow and the results are well interpreted. The methods section provides good detail and data are accessible in public databases.

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Very good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
No: Not applicable

Author response to reviewers to Version 1

Response to Reviewers' comments:

Reviewer 1

1. Where the authors discuss the functions of biosynthetic genes identified (lines 90-116), can any hypothesis be made regarding which genes or BGCs may be responsible for the antifungal activity of the strain observed in previous studies? While the characterisation of these genes/BGCs is clearly beyond the scope of this short communication, some discussion of this would add to the manuscript. Some short information about that was incorporated in lines 102 – 104 and 118 – 120. Besides that, a disclaimer was add in the manuscript conclusion, in lines 225 – 227.

2. Table 1: The authors present analysis of the BGCs predicted to be present in the genome from AntiSMASH analysis and some key proteins present. Could any more detail regarding most similar BGCs in the database be added here to give more ideas about the biosynthetic potential of the strain? We made some changes at this table and included some details about similar BGCs in antiSMAH in lines 132 – 135.

3. Some revisions of the English could be applied through throughout the manuscript, with special attention to tenses and grammar. Some examples include:

a. Line 25: 'with emphasize in…' should be 'with emphasis in' Done it.

b. Line 85-86: 'it was detected' should be 'genes were detected' Done it.

c. Line 91: 'One of this is' should be 'One of these is' Done it.

4. Line 61: Typo 'estreptomycete' It was corrected to “streptomycete”

Reviewer 2

1.Methodological rigour

Without access to the genome sequence itself, the overall quality of the genome assembly cannot be determined. I would therefore strongly recommend using software such as QUAST to show the number of contigs in your assembly, NC50 values etc. I feel this is particularly relevant given that Illumina sequencing alone is used here- without long read sequencing to bridge the gaps. You may wish to use software such as bandage- available on galaxy- to to display your contigs. This would also allow for the reader to guage how accurate the antismash predictions are likely to be. I feel this information is essential for the publishing of this paper. Information about the contigs methodology was added in lines 64 – 65 and information about contigs and N50 in lines 85 – 86.

2.Presentation of results

The phylogenetic trees which are presented show the species which are most closely related in a concise and effective way. However, although it appears 102 strains were used in the Neighbor-joining phenetic tree only a small number are shown. I would recommended using something like figtree to download the .newick file and present the whole tree to determine where the strain lies. You should also describe the software used for tree generation and display in your figure legend.The figure with all 102 genomes is in the Supplementary Material, together with a list of all the species that this 102 genomes comprise. We made corrections in the figure legend in line 201.

3.How the style and organization of the paper communicates and represents key findings

The organisation of the paper communicates the findings well.

4.Literature analysis or discussion

I feel there should be more discussion/comment regarding which gene cluster is likely to produce the anti-fungal compound of interest. A small figure showing the genes in some clusters may also be relevant. You may also wish to discuss the fact your strain doesn't appear to encode some very typical Streptomyces natural products such as hopene or geosmin.A small figure was included in Supplementary Material, we write about that in lines 132 – 135. Hopene and geosmin were discussed in lines 139 – 142.

In Line 207: You write that you cannot define this as a species due to overlapping 16S sequencing. Could you please clarify that you mean the species cannot be described genomically- but chemotaxinomic assays can be used in future define the species boundaries. Perhaps a general comment about speciation in bacteria being complex may help, with reference to average nucleotide identity being used as a broad technique. We talked about the difficult in lines 42 – 46 to introduce the subject and we added more information and discuss in lines 219 – 225.

5. Any other relevant comments

This paper needs to be thoroughly proof read for grammatical errors before re-submission. I have highlighted some below along with some other minor issues.

Line 39: Change 'finding limitations' to 'limited'Done it.

Line 26: Changed analysis to analyses. Define ST before using the abbreviation.Done it.

Line 30: Change 'who' to 'which'Done it.

Method:Could we have more information about the source if possible? Species of tomato, geographical location of the tomato plant etc. Information added in line 53 – 54.

Line 61: estreptomycete? Corrected to “streptomycete”

VERSION 1

Editor recommendation and comments

https://doi.org/10.1099/acmi.0.000466.v1.5

John Munnoch; University of Strathclyde, SIPBS, UNITED KINGDOM, Glasgow

Date report received: 18 July 2022

Recommendation: Minor Amendment

Comments: This study would be a valuable contribution to the existing literature. The reviewers have highlighted minor concerns with the work presented. Please ensure that you address their comments.

Reviewer 2 recommendation and comments

https://doi.org/10.1099/acmi.0.000466.v1.3

Rebecca McHugh; University of Glasgow Institute of Infection Immunity and Inflammation, Bacteriology, UNITED KINGDOM
https://orcid.org/0000-0003-4664-6125

Date report received: 14 July 2022
Recommendation: Minor Amendment

Comments: The authors have presented this manuscript which describes the whole genome sequencing of Streptomyces sp. 6(4), and the mining of this genome for specialised metabolite gene clusters. In this study they show that Streptomyces strain in question encodes 11 special metabolite gene clusters which include typical Streptomyces BGCs such as melanin, as well as T1PKSs and NRPS gene clusters. Streptomyces genome sequencing is often complex and you have been successful in obtaining a genome sequence which is sufficient to help you define the BGC responsible for antifungal production. 1. Methodological rigour Without access to the genome sequence itself, the overall quality of the genome assembly cannot be determined. I would therefore strongly recommend using software such as QUAST to show the number of contigs in your assembly, NC50 values etc. I feel this is particularly relevant given that Illumina sequencing alone is used here- without long read sequencing to bridge the gaps. You may wish to use software such as bandage- available on galaxy- to to display your contigs. This would also allow for the reader to guage how accurate the antismash predictions are likely to be. I feel this information is essential for the publishing of this paper 2. Presentation of results The phylogenetic trees which are presented show the species which are most closely related in a concise and effective way. However, although it appears 102 strains were used in the Neighbor-joining phenetic tree only a small number are shown. I would recommended using something like figtree to download the .newick file and present the whole tree to determine where the strain lies. You should also describe the software used for tree generation and display in your figure legend. 3. How the style and organization of the paper communicates and represents key findings The organisation of the paper communicates the findings well. 4. Literature analysis or discussion I feel there should be more discussion/comment regarding which gene cluster is likely to produce the anti-fungal compound of interest. A small figure showing the genes in some clusters may also be relevant. You may also wish to discuss the fact your strain doesn't appear to encode some very typical Streptomyces natural products such as hopene or geosmin. In Line 207: You write that you cannot define this as a species due to overlapping 16S sequencing. Could you please clarify that you mean the species cannot be described genomically- but chemotaxinomic assays can be used in future define the species boundaries. Perhaps a general comment about speciation in bacteria being complex may help, with reference to average nucleotide identity being used as a broad technique. 5. Any other relevant comments This paper needs to be thoroughly proof read for grammatical errors before re-submission. I have highlighted some below along with some other minor issues. Line 39: Change 'finding limitations' to 'limited' Line 26: Changed analysis to analyses. Define ST before using the abbreviation. Line 30: Change 'who' to 'which' Method: Could we have more information about the source if possible? Species of tomato, geographical location of the tomato plant etc. Line 61: estreptomycete?

Please rate the manuscript for methodological rigour
Satisfactory

Please rate the quality of the presentation and structure of the manuscript
Good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
Yes

Reviewer 1 recommendation and comments

https://doi.org/10.1099/acmi.0.000466.v1.4

Rebecca Devine; John Innes Centre, UNITED KINGDOM
https://orcid.org/0000-0003-0008-2184

Date report received: 13 July 2022
Recommendation: Minor Amendment

Comments: The manuscript presented describes the full genome sequence of Streptomyces sp. 6(4), a strain previously isolated from tomato root that shows antifungal activity against the plant-pathogen, Bipolaris sorokiniana. The MiSeq Illumina platform was used to sequence the genomic DNA and the genes present have been annotated using several methods. Phylogenetic analysis has been conducted to identify the species of the isolate and AntiSMASH analysis to predict biosynthetic genes and gene clusters present. The manuscript is generally well presented. The workflow is easy to follow and the results are well interpreted. The methods section provides good detail and data are accessible in public databases. My specific recommendations are: 1. Where the authors discuss the functions of biosynthetic genes identified (lines 90-116), can any hypothesis be made regarding which genes or BGCs may be responsible for the antifungal activity of the strain observed in previous studies? While the characterisation of these genes/BGCs is clearly beyond the scope of this short communication, some discussion of this would add to the manuscript. 2. Table 1: The authors present analysis of the BGCs predicted to be present in the genome from AntiSMASH analysis and some key proteins present. Could any more detail regarding most similar BGCs in the database be added here to give more ideas about the biosynthetic potential of the strain? 3. Some revisions of the English could be applied through throughout the manuscript, with special attention to tenses and grammar. Some examples include: a. Line 25: 'with emphasize in…' should be 'with emphasis in' b. Line 85-86: 'it was detected' should be 'genes were detected' c. Line 91: 'One of this is' should be 'One of these is' 4. Line 61: Typo 'estreptomycete'

Please rate the manuscript for methodological rigour
Good

Please rate the quality of the presentation and structure of the manuscript
Very good

To what extent are the conclusions supported by the data?
Strongly support

Do you have any concerns of possible image manipulation, plagiarism or any other unethical practices?
No

Is there a potential financial or other conflict of interest between yourself and the author(s)?
No

If this manuscript involves human and/or animal work, have the subjects been treated in an ethical manner and the authors complied with the appropriate guidelines?
No: Not applicable

SciScore report

https://doi.org/10.1099/acmi.0.000466.v1.1

iThenticate report

https://doi.org/10.1099/acmi.0.000466.v1.2