

Genome sequence of the aurodox-producing bacterium *Streptomyces goldiniensis* ATCC 21386

Rebecca E. McHugh^{1,2}, John T. Munnoch¹, Andrew J. Roe² and Paul A. Hoskisson^{1,*}

Abstract

We report the genome sequence of *Streptomyces goldiniensis* ATCC 21386, a strain which produces the anti-bacterial and anti-virulence polyketide, aurodox. The genome of *S. goldiniensis* ATCC 21386 was sequenced using a multiplatform hybrid approach, revealing a linear genome of ~10 Mbp with a G+C content of 71%. The genome sequence revealed 36 putative biosynthetic gene clusters (BGCs), including a large region of 271 Kbp that was rich in biosynthetic capability. The genome sequence is deposited in DDBJ/EMBL/GenBank with the accession number PRJNA602141.

INTRODUCTION

Isolated from soil collected in Bermuda in 1973, a novel strain of *Streptomyces* was found to produce the anti-Streptococcal natural product (X-5108=aurodox) and was informally named *S. goldiniensis var. goldiniensis* [1]. Despite the strain being named *S. goldiniensis var. goldiniensis* by Berger *et al.* [1], there is no formal description of this strain in the literature and it does not appear in the List of Prokaryotic names with standing in nomenclature [2].

Originally identified for its antibacterial activity, aurodox [1] has also found utility as a widely used growth promoting compound in poultry [3]. More recently it has attracted attention for its anti-virulence activity, blocking virulence by inhibition of the Type III Secretion System (T3SS) in Enterohaemorrhagic *Escherichia coli* (EHEC) [4]. Subsequent work this inhibitory effect is mediated through the down-regulation of the Type III Secretion System master regulator Ler [5].

Whilst *S. goldiniensis var. goldiniensis* is known for the production of aurodox, it is a common feature of *Streptomyces* genomes to encode many more secondary metabolites than can be observed during laboratory culture [6]. The vast repository of natural product biosynthetic gene clusters (BGCs) contained within the genomes of *Streptomyces* species means that genome mining has become the mainstay of researchers looking to prioritize BGCs for further study [6, 7]. Moreover, the increasing amount of genome sequence data of natural product producing strains facilitates evolutionary studies of biosynthesis and can inform on synthetic biology strategies for developing novel molecules and improving production of current molecules [6–8]. Here we describe the multi-platform genome sequencing of *Streptomyces goldiniensis* ATCC 21386 which is deposited in DDBJ/EMBL/GenBank with the accession number PRJNA602141.

METHODS

Whole genome sequencing *Streptomyces goldiniensis*.

Genomic DNA was extracted according to Kieser *et al.*, [9] from cultures grown in GYM medium (DSMZ medium 65; www.dsmz.de). Nanopore sequencing was performed using the Nanopore 1D ligation protocol with MinION SPOT ON MK1 R9 flow cells. Raw data was converted using MinKnow base calling software. Illumina platform data was provided by Microbes NG (Birmingham, UK) from the HiSeq 2500 sequencing platform. Reads were trimmed using Trimmomatic 0.30 [10] with quality cut off of Q15. PacBio sequencing was provided by (Nu-omics, University of Northumbria, UK) using the Sequel

Received 15 February 2022; Accepted 08 April 2022

Author affiliations: ¹Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde, 161 Cathedral Street, Glasgow, G4 0RE, UK;

²Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, G12 8TA, UK.

***Correspondence:** Paul A. Hoskisson, paul.hoskisson@strath.ac.uk

Keywords: Streptomyces; Aurodox; elfamycin; antimicrobial; genome.

Genome assemblies can be found Genbank Bioproject PRJNA602141. Genome Sequencing read raw data can be found at the Sequence Read Archive (SRA) and Figshare <https://dataview.ncbi.nlm.nih.gov/object/SRR18547533>, <https://dataview.ncbi.nlm.nih.gov/object/SRR18547532>, <https://doi.org/10.6084/m9.figshare.19487693.v2>.

000358 © 2022 The Authors



This is an open-access article distributed under the terms of the Creative Commons Attribution License. This article was made open access via a Publish and Read agreement between the Microbiology Society and the corresponding author's institution.

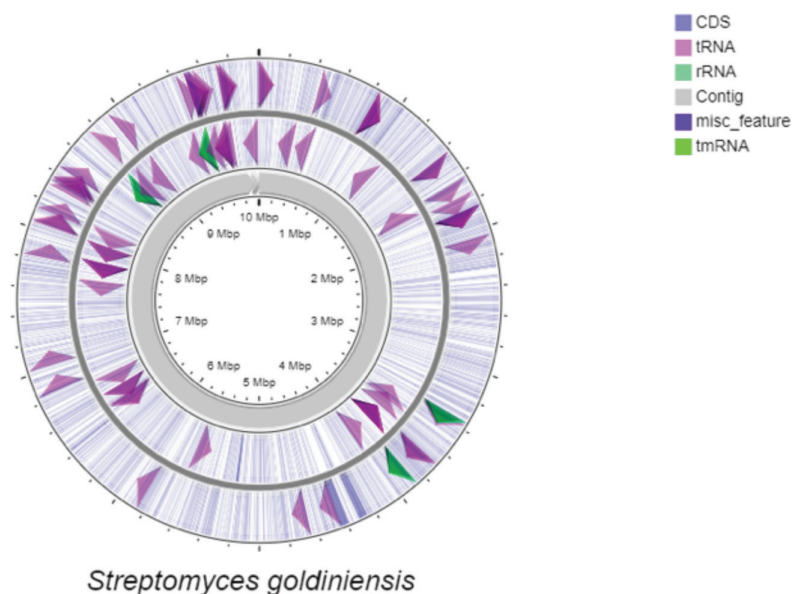


Fig. 1. Circular representation of the linear *Streptomyces goldiniensis* genome generated by CGviewer [19]. Map was built from a .gbk file generated by Prokka. The outer most circle represents coding sequences on the major strand, the inner circle represents coding sequences on the minor strand. Grey arrows in the centre of the figure depict the lengths of the contigs in the final assembly.

instrument with contigs assembled in HGAP4. Raw read data (Illumina, PacBio and ONT) is deposited in the Sequence Read Archive (SRA) at <https://dataview.ncbi.nlm.nih.gov/object/SRR18547533>, <https://dataview.ncbi.nlm.nih.gov/object/SRR18547532> and Figshare <https://doi.org/10.6084/m9.figshare.19487693.v2>

The SPAdes platform was used to create a combined assembly using data from all three technologies. This allowed for the use of a k-mer dependent approach whilst using the initial PacBio assembly as ‘trusted contigs’ [11]. AutoMLST [12] was used to identify *S. bottropensis* ATCC 25435 (Taxonomy ID: 1054862) as the closest neighbour for scaffold-based assembly using MeDuSa [13] and quality analysis performed using QUAST [14]. Prokka was used to annotate the genome of *Streptomyces goldiniensis* [15] and is available on Genbank (Bioproject PRJNA602141). Identification of biosynthetic gene clusters was performed using the antiSMASH pipeline (bacterial version 5.0.0) [16]. The position of the putative aurodox BGC within the larger 273 Kbp ‘supercluster’ was confirmed via PCR. The following oligonucleotide primer sequences were used; clusterposcheckAF 5'-CCAGACGCAGGTCCGCTTCGGACG-3'; clusterposcheckAR 5'-CCATCGTGGGGATCGCAG-3'; clusterposcheckBF 5'-AGGATGTTCCAGTCCGCTCTCACTCCG-3'; clusterposcheckBR 5'-CGAGGTCCCGCCGGCATGTGGA-3'.

PCR products were visualised using agarose gel-electrophoresis, and sequence specificity was confirmed by band excision followed by Sanger sequencing provided by Eurofins genomics (Luxembourg).

RESULTS AND DISCUSSION

Genome features of *Streptomyces goldiniensis* ATCC 21386

The linear genome of *S. goldiniensis* ATCC 21386 was sequenced using a hybrid-approach of Illumina, PacBio and Oxford Nanopore to generate a high-quality draft genome (Genbank Bioproject PRJNA602141). Using a combined assembly approach with SPAdes [11] the data from all three platforms allowed the overall genome size to be estimated at 10005022 bp, in nine contigs, with an N50 of 9950726 bp. The draft genome of *S. goldiniensis* is predicted to have a total of 9925 protein coding genes, along with 81 tRNAs and five rRNA operons (Fig. 1). Absence of genes encoding plasmid replication machinery (Par proteins) on the eight minor contigs suggest that they do not represent plasmids. Pulse-field gel electrophoresis of total DNA extractions from *S. goldiniensis* also indicates the absence of plasmids in this strain.

The *S. goldiniensis* genome is rich in natural product biosynthetic gene clusters

Streptomyces bacteria are renowned for their ability to synthesise a wide range of natural products, many of which have found utility in human medicine [7]. Despite many strains being identified through the production of a single metabolite, the genomes of *Streptomyces* often encode a number of additional biosynthetic gene clusters (BGCs) that are not expressed under

Table 1. Predicted Biosynthetic Gene Clusters of *S. goldiniensis* using antiSMASH [17]

Region	Type	Region	Most similar	% similarity
1	Type II PKS	32,278–104 355	Spore pigment	83
3	Lanthipeptide, lasso peptide	486,431–514 192	Citrulassin E (ripp)	100
4	Siderophore	2,159,777–2 170 184	N/A	0
5	Terpene	2,487,747–2 508 882	N/A	4
6	NRPS, tfuA related	2,543,021–2 606 769	Diisonitrile antibiotic SF2768	66
7	Bacteriocin	2,705,875–2 716 826	N/A	0
8	Terpene	2,777,350–2 799 499	Geosmin	100
9	NRPS-like	2,818,277–2 858 728	S56P1	11
10	NRPS-like, siderophore	3,085,338–3 126 119	Paulomycin	13
11	NRPS-like	3,541,716–3 582 164	Octacosamicin	12
12	Terpene	3,625,827–3 643 980	N/A	0
13	Indole	4,158,688–4 181 228	Terfestatin	28
14	bacteriocin, bottromycin, NRPS, transAT-PKS, Type I PKS, hglE-KS, lasso peptide	4,213,370–4 484 508	Kirromycin/bottromycin	40
15	NRPS-like	4,599,349–4 657 342	Echosides	17
16	Terpene	4,783,720–4 802 511	N/A	0
17	Linardin	5,046,397–5 067 017	Pentostatine	17
18	Siderophore	5,085,773–5 100 346	N/A	0
19	NRPS, melanin	5,167,798–5 232 264	Scabichelin	100
20	Terpene	5,257,095–5 277 644	N/A	0
21	NRPS, Type I PKS	5,284,180–5 346 580	N/A	4
22	Terpene	5,480,336–5 505 711	Isorenieratene	100
23	Type II PKS	5,770,343–5 856 353	Fluostatins	60
24	Lanthipeptide, bacteriocin	6,153,412–6 179 099	Informatipeptin	100
25	NRPS-like, Type I PKS	6,410,852–6 457 973	Marineosin	45
26	Terpene	6,507,863–6 532 462	Hopene	100
27	Ectoine	6,852,064–6 861 735	Ectoine	100
28	Ladderane	6,937,157–6 978 500	Colabomycin	11
29	Type III PKS, Terpene	7,764,486–7 805 661	Merochlorin	12
30	Melanin	8,189,939–8 200 406	Melanin	80
31	Siderophore	8,344,177–8 355 949	Desferrioxamine A	83
32	Ectoine	8,471,171–8 481 599	Ectoine	75
33	Terpene	9,170,031–9 191 797	N/A	0
34	Bacteriocin	9,838,044–9 848 331	N/A	0
35	Terpene	9,968,421–10 005 017	N/A	0

laboratory conditions, the so called ‘silent BGCs’ [6]. Using the antiSMASH pipeline [16] the genome of *S. goldiniensis* ATCC 21386 was mined. A total of 36 putative BGCs were identified (Table 1), including five putative polyketide synthase (PKS) containing BGCs, eight non-ribosomal peptide synthase (NRPS) BGCs and nine putative terpene BGCs. The genome encodes BGCs that are highly conserved in the genomes of *Streptomyces* species such as geosmin, desferrioxamine and melanin [7].

One region of the *S. goldiniensis* genome was found to be particularly rich in BGCs (position 4213370–4484508; 271 kb), encoding a putative bottromycin A2-like molecule, an 87 kb region that possesses genes likely to encode a hybrid PKS/NRPS,

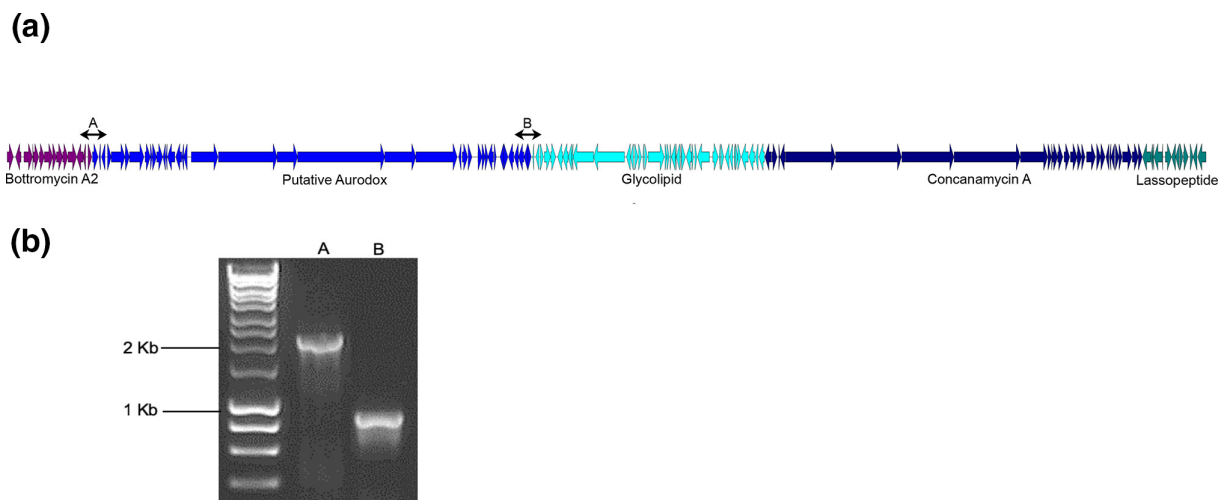


Fig. 2. (a): Representation of the 271 kbp putative aurodox supercluster. Gene cluster boundaries were predicted by AntiSMASH and schematic diagram was generated by clinker. From left to right, arrows coloured purple represent Bottromycin A2-associated genes, blue arrows represent the putative aurodox BGC, cyan arrows represent predicted glycolipid-synthase genes, navy blue arrows represent concanamycin A-associated genes and green arrows indicate putative genes involved in lasso peptide biosynthesis. (b): Primer pair binding in lasso peptide biosynthesis. (b): Primer pair binding sites are indicated by 'a' and 'b' on the diagram and their corresponding PCR products are shown on the gel with the lanes labelled accordingly.

highly similar to the kirromycin BGC [17] which has recently been shown to encode aurodox [18]. Immediately downstream of the aurodox BGC is a gene cluster with homology to glycolipid synthase-like PKS containing BGCs, a cluster which is 100% identical to the macrolide concanamycin A and a putative lasso peptide-encoding gene cluster (Table 1, Fig. 2). To confirm that this BGC-rich region around the aurodox BGC was indeed a supercluster, rather than an artefact of genome assembly, primer pairs were designed that spanned the junction between the cluster upstream (putative bottromycin A2) and downstream (glycolipid synthase-like PKS containing BGC) of the aurodox cluster. PCR and sequencing of the regions spanning the BGCs confirmed the organisation of the gene clusters around aurodox (Fig. 2).

CONCLUSIONS

The genome of *S. goldiniensis* ATCC 21386 was sequenced using a hybrid approach to yield a high-quality draft genome with ~99% of the genome on a single contig through a k-mer dependant assembly using SPAdes [11], followed by a scaffold-based final assembly with MeDuSa [13]. This allowed prediction of a biosynthetic gene supercluster to be identified from the main genome contig and the organisation of the supercluster to be confirmed by PCR. This genome sequence provides a springboard for further study of this strain and a basis for a formal taxonomic description of *S. goldiniensis* var. *goldiniensis* ATCC 21386.

Funding information

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

Acknowledgements

We would like to thank the University of Strathclyde and University of Glasgow for jointly funding the PhD of REM. PAH would also like to acknowledge funding from BBSRC (BB/T001038/1 and BB/T004126/1) and the Royal Academy of Engineering Research Chair Scheme for long term personal research support (RCSRF2021\11\15). AJR and PAH would like to acknowledge funding from MRC (MR/V011499/1). We would also like to thank Professor Iain S. Hunter (University of Strathclyde) and Professor Matt Hutchings (John Innes Centre) for helpful discussions.

Conflicts of interest

The authors declare that there are no conflicts of interest.

References

- Berger J, Lehr HH, Teitel S, Maehr H, Grunberg E. A New antibiotic x-5108 of streptomycetes origin. *J Antibiot* 1973;26:15–22.
- Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List of prokaryotic names with standing in nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Microbiol* 2020;70:5607–5612.
- Marusich WL, Ogrinz EF, Mitrovic M. A New Antibiotic, X-5108, for Improved Growth and Feed Conversion in Poultry. *Poult Sci* 1974;53:936–945.
- Kimura K, Iwatsuki M, Nagai T, Matsumoto A, Takahashi Y, et al. A small-molecule inhibitor of the bacterial type III secretion system protects against *in vivo* infection with *Citrobacter rodentium*. *J Antibiot (Tokyo)* 2011;64:197–203.

5. McHugh RE, O'Boyle N, Connolly JPR, Hoskisson PA, Roe AJ. Characterization of the Mode of Action of Aurodox, a Type III Secretion System Inhibitor from *Streptomyces goldiniensis*. *Infect Immun* 2019;87:e00595-18.
6. Hoskisson PA, Seipke RF. Cryptic or silent? the known unknowns, unknown knowns, and unknown unknowns of secondary metabolism. *mBio* 2020;11:e02642-20.
7. Chevrette MG, Gutiérrez-García K, Selem-Mojica N, Aguilar-Martínez C, Yañez-Olvera A, *et al.* Evolutionary dynamics of natural product biosynthesis in bacteria. *Nat Prod Rep* 2020;37:566–599.
8. Fernández-Martínez LT, Hoskisson PA. Expanding, integrating, sensing and responding: the role of primary metabolism in specialised metabolite production. *Curr Opin Microbiol* 2019;51:16–21.
9. Kieser T, Bibb MJ, Buttner MJ, Chater KF, Hopwood D. *Practical Streptomyces Genetics*. John Innes Foundation, 2000.
10. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 2014;30:2114–2120.
11. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, *et al.* SPAdes: A new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
12. Alanjary M, Steinke K, Ziemert N. AutoMLST: an automated web server for generating multi-locus species trees highlighting natural product potential. *Nucleic Acids Res* 2019;47:W276–W282.
13. Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, *et al.* MeDuSa: a multi-draft based scaffolder. *Bioinformatics* 2015;31:2443–2451.
14. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 2013;29:1072–1075.
15. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30:2068–2069.
16. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, *et al.* antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 2019;47:W81–W87.
17. Weber T, Laiple KJ, Pross EK, Textor A, Grond S, *et al.* Molecular analysis of the kirromycin biosynthetic gene cluster revealed beta-alanine as precursor of the pyridone moiety. *Chem Biol* 2008;15:175–188.
18. McHugh RE, Munnoch JT, Braes RE, McKean IJW, Giard JM, *et al.* Biosynthesis of aurodox, a Type III secretion system inhibitor from *Streptomyces goldiniensis*. *bioRxiv* [Preprint]; 2022.02.14.480342 2022. DOI: 10.1101/2022.02.14.480342.
19. Stothard P, Wishart DS. Circular genome visualization and exploration using CGView. *Bioinformatics* 2005;21:537–539.

Five reasons to publish your next article with a Microbiology Society journal

1. The Microbiology Society is a not-for-profit organization.
2. We offer fast and rigorous peer review – average time to first decision is 4–6 weeks.
3. Our journals have a global readership with subscriptions held in research institutions around the world.
4. 80% of our authors rate our submission process as 'excellent' or 'very good'.
5. Your article will be published on an interactive journal platform with advanced metrics.

Find out more and submit your article at microbiologyresearch.org.