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

Rebecca E. McHugh[1], John T. Munnoch[1], Andrew J Roe[2] and Paul A. Hoskisson[1*]

[1]Strathclyde Institute of Pharmacy and Biomedical Science, University of Strathclyde, 161 Cathedral Street, Glasgow, G4 0RE. UK. Tel. +44 (0)141 548 2819 Fax +44 (0)141 548 4124 Email: Paul.hoskisson@strath.ac.uk

[2*]Institute of Infection, Immunity and Inflammation, University of Glasgow, Glasgow, G12 8TA, UK

[*]Corresponding author

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 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 observed during laboratory culture [6]. The vast repository of natural product biosynthetic gene clusters (BGCs) contained within the genomes of *Streptomyces* 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 [7–9]. 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.

52 **Results & Discussion**

53

54 **Genome features of *Streptomyces goldiniensis* ATCC 21386**

55 The linear genome of *S. goldiniensis* ATCC 21386 was sequenced using a hybrid-
56 approach of Illumina, PacBio and Oxford Nanopore to generate a high-quality draft
57 genome (Genbank Bioproject PRJNA602141). Using a combined assembly approach with
58 SPAdes [10] the data from all three platforms allowed the overall genome size to be
59 estimated at 10,005,022 bp, in nine contigs, with an N50 of 9,950,726 bp. The draft
60 genome of *S. goldiniensis* is predicted to have a total of 9925 protein coding genes, along
61 with 81 tRNAs and five rRNA operons (**Figure 1**). Absence of genes encoding plasmid
62 replication machinery (Par proteins) on the eight minor contigs suggest that they do not
63 represent plasmids. Pulse-filed Gel electrophoresis of total DNA extractions from *S.*
64 *goldiniensis* also indicates the absence of plasmids in this strain.

65

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

67

68 *Streptomyces* bacteria are renowned for their ability to synthesis a wide range of natural
69 products, many of which have found utility in human medicine [7]. Despite many strains
70 being identified through the production of a single metabolite, the genomes of
71 *Streptomyces* often encode a number of additional biosynthetic gene clusters (BGCs) that
72 are not expressed under laboratory conditions, the so called 'silent BGCs' [6]. Using the
73 antiSMASH pipeline [11] the genome of *S. goldiniensis* ATCC 21386 was mined. A total
74 of 36 putative BGCs were identified (**Table 1**), including five putative polyketide synthase
75 (PKS) containing BGCs, eight non-ribosomal peptide synthase (NRPS) BGCs and nine
76 putative terpene BGCs. The genome encodes BGCs that are highly conserved in the
77 genomes of *Streptomyces* species such as geosmin, desferrioxamine and melanin [7].

78

79 One region of the *S. goldiniensis* genome was found to be particularly rich in BGCs
80 (position 4,213,370 - 4,484,508; 271 kb), encoding a putative Bottromycin A2-like
81 molecule, an 87 kb region that possesses genes likely to encode a hybrid PKS/NRPS,
82 highly similar to the kirromycin BGC [12] which has recently been shown to encode
83 aurodox (McHugh *et al., in prep.*). Immediately downstream of the aurodox BGC is a gene
84 cluster with homology to glycolipid synthase-like PKS containing BGCs, a cluster which is

100 % identical to the macrolide concanamycin A and a putative lassopeptide-encoding gene cluster (**Table 1; Figure 2**). To confirm that the 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 (**Figure 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 [10], followed by a scaffold-based final assembly with [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.

Materials and Methods

Whole Genome Sequencing *Streptomyces goldiniensis.*

Genomic DNA was extracted according to Kieser *et al,*[14] 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[15] with quality cut off of Q15. PacBio sequencing was provided by (Nu-omics, University of Northumbria, UK) using the Sequel instrument with contigs assembled in HGAP4.

The genome was assembled using SPAdes with combined data from all three platforms [10]. AutoMLST (38) 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 [16]. Prokka was used to annotate the genome of *Streptomyces goldiniensis* [17] and is available on Genbank (Bioproject PRJNA602141). Identification of biosynthetic gene clusters was performed using the antiSMASH pipeline (bacterial version 5.0.0) [11]. 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'- AGGATGTTCCAGTCGGCTCTCACTCCG-3';

clusterposcheckBR 5'-CGAGGTCGCCCGGCATGTGGA-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).

Funding Information

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).

Funding Statement

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

Acknowledgements

We would also like to thank Professor Iain S. Hunter (University of Strathclyde) and Professor Matt Hutchings (John Innes Centre) for helpful discussions.

146 **Figure Legends**

147

148 **Figure 1:** Circular representation of the linear *Streptomyces goldiniensis* genome
149 generate by CGviewer [18]. Map was built from a .gbk file generated by Prokka. The outer
150 most circle represents coding sequences on the major strand, the inner circle represents
151 coding sequences on the minor strand. Grey arrows in the centre of the figure depict the
152 lengths of the contigs in the final assembly.

153

154 **Figure 2: A: Representation of the 271 kbp putative aurodox supercluster.** Gene
155 cluster boundaries were predicted by AntiSMASH and schematic diagram was generated
156 by clinker. From left to right, arrows coloured purple represent Bottromycin A2-associated
157 genes, blue arrows represent the putative aurodox BGC, cyan arrows represent predicted
158 glycolipid-synthase genes, navy blue arrows represent concanamycin A-associated genes
159 and green arrows indicate putative genes involved in lassopeptide biosynthesis. **B:** Primer
160 pair binding sites are indicated by 'A' and 'B' on the diagram and their corresponding PCR
161 products are shown on the gel with the lanes labelled accordingly.

162

References

1. **Berger J, Lehr HH, Teitel S, Maehr H, Grunberg E**. A new antibiotic X-5108 of Streptomyces origin. *Journal of Antibiotics* 1973;26:15–22.

2. **Parte AC, Carbasse JS, Meier-Kolthoff JP, Reimer LC, Göker M**. List of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the DSMZ. *Int J Syst Evol Micr*. Epub ahead of print 2020. DOI: 10.1099/ijsem.0.004332.

3. **Marusich WL, Ogrinz EF, Mitrovic M**. A New Antibiotic, X-5108, for Improved Growth and Feed Conversion in Poultry. *Poultry Sci* 1974;53:936–945.

4. **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 Antibiotics* 2010;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*;87. Epub ahead of print 2018. DOI: 10.1128/iai.00595-18.

6. **Hoskisson PA, Seipke RF**. Cryptic or Silent? The Known Unknowns, Unknown Knowns, and Unknown Unknowns of Secondary Metabolism. *Mbio*;11. Epub ahead of print 2020. DOI: 10.1128/mbio.02642-20.

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*. Epub ahead of print 2019. DOI: 10.1039/c9np00048h.

8. **Chevrette MG, Hoskisson PA, Barona-Gómez F**. Enzyme Evolution in Secondary Metabolism. *Reference Modul Chem Mol Sci Chem Eng*. Epub ahead of print 2019. DOI: 10.1016/b978-0-12-409547-2.14712-2.

9. **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.

10. **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.

11. **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.

12. **Weber T, Laiple KJ, Pross EK, Textor A, Grond S, *et al.*** Molecular Analysis of the Kirromycin Biosynthetic Gene Cluster Revealed β-Alanine as Precursor of the Pyridone

199 Moiety. *Chemistry & Biology*;15. Epub ahead of print 2008. DOI:
200 10.1016/j.chembiol.2007.12.009.

201 13. **Bosi E, Donati B, Galardini M, Brunetti S, Sagot M-F, *et al.*** MeDuSa: a multi-draft
202 based scaffolder. *Bioinformatics* 2015;31:2443–2451.

203 14. **Kieser T, Bibb M, Buttner M, Chater K, Hopwood, *et al.*** *Practical Streptomyces*
204 *Genetics*. 2000.

205 15. **Bolger AM, Lohse M, Usadel B**. Trimmomatic: a flexible trimmer for Illumina
206 sequence data. *Bioinformatics* 2014;30:2114–2120.

207 16. **Gurevich A, Saveliev V, Vyahhi N, Tesler G**. QUAST: quality assessment tool for
208 genome assemblies. *Bioinformatics* 2013;29:1072–1075.

209 17. **Seemann T**. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*
210 2014;30:2068–2069.

211 18. **Stothard P, Wishart DS**. Circular genome visualization and exploration using
212 CGView. *Bioinformatics* 2005;21:537–539.

213

214

215

216 **Table 1:** Summary of the predicted Biosynthetic Gene Clusters of *S. goldiniensis* using

217 antiSMASH *[11]*.

218

Region	Type	Region	Most similar	% similarity
1	Type II PKS	32,278 - 104,355	Spore pigment	83
3	Lanthipeptide, lassopeptide	486,431 - 514,192	Citrulassin E (ripp)	100
4	Siderophore	2,159,777 - 2,170,184	NA	0
5	Terpene	2,487,747 - 2,508,882	Lipopeptide	4
6	NRPS, tfuA related	2,543,021 - 2,606,769	Diisonitrile antibiotic SF2768	66
7	Bacteriocin	2,705,875 - 2,716,826	NA	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	NA	0
13	Indole	4,158,688 - 4,181,228	Terfestatin	28
14	bacteriocin,bottromycin,NRPS,transAT-PKS, Type I PKS, hglE-KS,lassopeptide	4,213,370 - 4,484,508	Kirromycin	40
15	NRPS-like	4,599,349 - 4,657,342	Echosides	17
16	Terpene	4,783,720 - 4,802,511	NA	0
17	Linardin	5,046,397 - 5,067,017	Pentostatine	17
18	Siderophore	5,085,773 - 5,100,346	NA	0
19	NRPS, melanin	5,167,798 - 5,232,264	Scabichelin	100
20	Terpene	5,257,095 - 5,277,644	NA	0
21	NRPS, Type I PKS	5,284,180 - 5,346,580	Herboxidiene	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	NA	0
34	Bacteriocin	9,838,044 - 9,848,331	NA	0
35	Terpene	9,968,421 - 10,005,017	NA	0

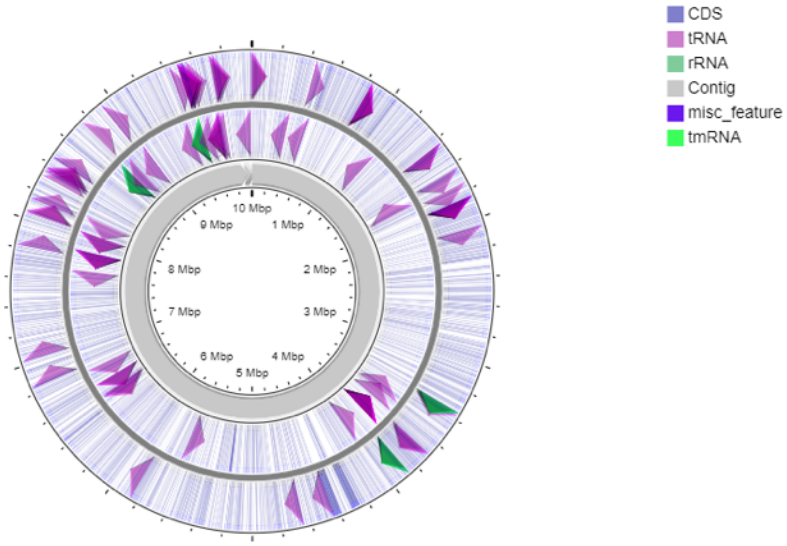
219

220

221 **Figure 1**

222

223



Streptomyces goldiniensis

224

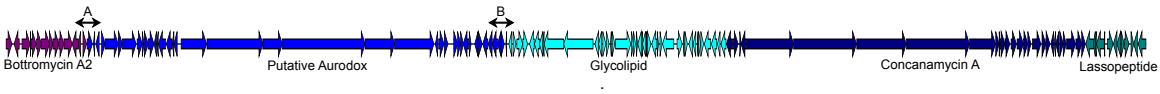
225

226

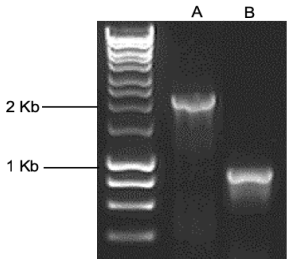
227 **Figure 2**

228

A



B



229

230