



Contents lists available at ScienceDirect

## Journal of Biotechnology

journal homepage: [www.elsevier.com/locate/jbiotec](http://www.elsevier.com/locate/jbiotec)

Short communication

Complete genome sequence of *Streptomyces formicae* KY5, the formicamycin producerNeil A. Holmes<sup>a,\*</sup>, Rebecca Devine<sup>a</sup>, Zhiwei Qin<sup>b</sup>, Ryan F. Seipke<sup>c</sup>, Barrie Wilkinson<sup>b</sup>, Matthew I. Hutchings<sup>a</sup><sup>a</sup> School of Biological Sciences, University of East Anglia, Norwich Research Park, Norwich, NR4 7TJ, United Kingdom<sup>b</sup> Department of Molecular Microbiology, John Innes Centre, Norwich Research Park, Norwich, NR4 7UH, United Kingdom<sup>c</sup> School of Molecular and Cellular Biology, Astbury Centre for Structural Molecular Biology, University of Leeds, Leeds, LS2 9JT, United Kingdom

## ARTICLE INFO

## Keywords:

*Streptomyces formicae* KY5*Tetraponera penzigi*

Genome sequence

Antibiotics

Secondary metabolites

## ABSTRACT

Here we report the complete genome of the new species *Streptomyces formicae* KY5 isolated from *Tetraponera* fungus growing ants. *S. formicae* was sequenced using the PacBio and 454 platforms to generate a single linear chromosome with terminal inverted repeats. Illumina MiSeq sequencing was used to correct base changes resulting from the high error rate associated with PacBio. The genome is 9.6 Mbps, has a GC content of 71.38% and contains 8162 protein coding sequences. Predictive analysis shows this strain encodes at least 45 gene clusters for the biosynthesis of secondary metabolites, including a type 2 polyketide synthase encoding cluster for the antibacterial formicamycins. *Streptomyces formicae* KY5 is a new, taxonomically distinct *Streptomyces* species and this complete genome sequence provides an important marker in the genus of *Streptomyces*.

*Streptomyces formicae* KY5 is a new species, isolated from the African plant ant *Tetraponera penzigi* (Qin et al., 2017; Seipke et al., 2013). These ants nest inside specialised hollow swellings called domatia in their host Acacia plants where they grow a fungus as food (Blatrix et al., 2012). In return for housing, they protect their host plants from large herbivores, including elephants (Palmer et al., 2008). *S. formicae* produces an unidentified antifungal compound, which is active against the multidrug resistant human pathogen *Lomentospora prolificans*, and a group of pentacyclic polyketides, called formicamycins, which have potent antibacterial activity against clinical MRSA and VRE isolates. *S. formicae* has biosynthetic gene clusters (BGCs) encoding for at least 45 additional natural products. It is amenable to genetic engineering using Cas9 mediated genome editing, which was used to remove the entire ~40 kbp formicamycin BGC (Qin et al., 2017).

High molecular weight genomic DNA was prepared using the salting out method (Kieser et al., 2000). Sequencing was performed at the Earlham Institute (Norwich, UK) using Pacific Biosciences (PacBio) RSII SMRT technology. Assembly using the HGAP2 pipeline gave one larger contig of 9.3Mb (unitig 9) and three smaller fragments of 179 kbp (unitig 1), 111 kbp (unitig 2) and 27 kbp (unitig 10). Roche 454 sequencing was also performed at the Earlham Institute, leading to 615 contigs (N50 31695 bp) using the Newbler

assembler v 2.3. Overlapping PacBio and 454 contigs were then aligned using BLAST (Camacho et al., 2009) and fragments were merged to form a single contig and extend the terminal inverted repeats (Fig. S1).

Illumina sequencing was then carried out at the DNA Sequencing Facility, Department of Biochemistry, University of Cambridge, UK, using TruSeq PCR-free and Nextera Mate Pair libraries and a MiSeq 600 sequencer. Reads were mapped to the single contig generated above, and variants were called using breseq software (Barrick et al., 2014). Errors in the single contig, which likely arose through the higher error rate of PacBio (as seen for the *S. leeuwenhoekii* genome (Gomez-Escribano et al., 2015)), were predicted as variants by breseq and manually corrected. In total 124 errors were corrected (Table S1), most errors were associated with runs of Gs or Cs, with 83 additions and 37 deletions of bases and 4 errors were associated with small repeat sequences. The resulting genome sequence was annotated by calling the open reading frames (ORFs) using Prodigal (Hyatt et al., 2010), and then further annotated with RAST (Aziz et al., 2008) (Fig. S2). The tRNA and rRNA genes were also predicted using RAST. The fully annotated genome sequence was submitted to NCBI Genbank and assigned the accession number CP022685. 454 contigs, the initial PacBio assembly, as well as PacBio and Illumina

**Abbreviations:** AT, acyl transferase; BGC, biosynthetic gene cluster; FAS, fatty acid synthase; NRPS, nonribosomal peptide synthetase; ORF, open reading frame; PKS, polyketide synthase; RiPP, ribosomally synthesized and post-translationally modified peptides; SAM, S-adenosylmethionine; TIR, terminal inverted repeats

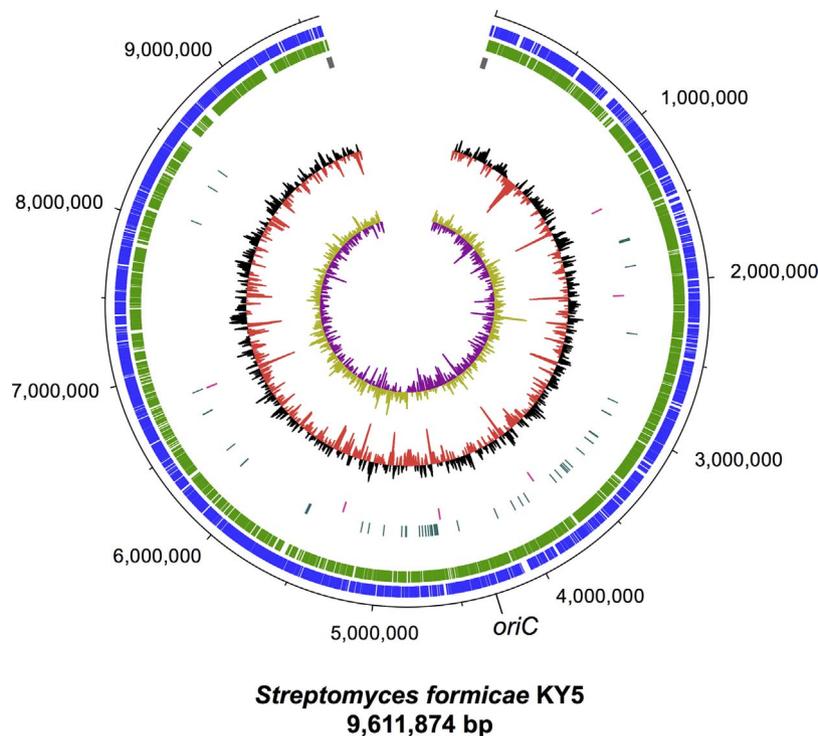
\* Corresponding author.

E-mail address: [n.holmes@uea.ac.uk](mailto:n.holmes@uea.ac.uk) (N.A. Holmes).<https://doi.org/10.1016/j.jbiotec.2017.11.011>

Received 31 October 2017; Received in revised form 17 November 2017; Accepted 25 November 2017

Available online 27 November 2017

0168-1656/© 2017 The Author(s). Published by Elsevier B.V. This is an open access article under the CC BY license (<http://creativecommons.org/licenses/by/4.0/>).



**Fig. 1.** Map of the *Streptomyces formicae* KY5 Genome. The outer scale is numbered in intervals of 0.5 Mbp. Circles 1 and 2 display the ORFs on the forward strand (blue) and reverse strand (green) respectively. Circle 3 displays the TIRs (grey). Circle 4 displays the tRNA genes (teal). Circle 5 displays the rRNA genes (pink). Circle 6 displays the GC percentage plot (black above average, red below average). Circle 7 displays the GC skew (lime green above average, purple below average). The origin of replication is marked *oriC*. The genome map was made using DNAPlotter v 10.2 (Carver et al., 2009). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

**Table 1**  
Genome features of *S. formicae* KY5.

Feature	Characteristic
Genome topology	Linear
Size	9,611,874 bp
GC content	71.38%
Open Reading Frames	8162
rRNA operons	6
tRNA genes	65
Secondary metabolite BGCs	34 (see Table 2)

reads (Mate Pair and TruSeq), were deposited at the NCBI Sequencing Reads Archive with accession number SRP117343.

The final polished assembly represents the complete genome sequence of *S. formicae* KY5. It comprises a single linear chromosome of 9,611,874 bp with a G + C content of 71.38% (Fig. 1 and Table 1). There is a centrally located origin of DNA replication *oriC* located between the *dnaA* and *dnaN* genes (4,801,506–4,802,428 bp). Terminal inverted repeats (TIRs) of 35,482 bp are present at the ends of the chromosome with each TIR ending in a terminal associated helicase *ttrA* gene (Huang et al., 2003). The genome contains 8162 predicted ORFs and encodes six rRNA operons and 65 tRNAs. Genes predicted to encode ORFs were given identifiers of “KY5” followed by 4 digits. The rRNAs and tRNAs were given identifiers “KY5\_rRNA” and “KY5\_tRNA” respectively, followed by 4 digits.

To assess the secondary metabolite biosynthetic potential, we used antiSMASH v 4.0 which predicts 34 BGCs, although manual inspection suggests several of these may be islands comprising multiple BGCs, taking the minimum number to 45 (Table 2). There is a single type 2 PKS gene cluster which is responsible for biosynthesis of the formicamycins (Qin et al., 2017), and BGCs encoding three type 1 PKSs, eleven NRPSs and a number of hybrid BGCs. Six BGCs encode for terpenes, including the antibiotic albaflavenone made by *Streptomyces*

**Table 2**

The 34 BGCs predicted by antiSMASH v 4.0 are numbered but manual inspection revealed that some of these are islands of two or more BGCs giving 45 BGCs in total. BGCs within the antiSMASH called clusters are annotated a, b, c, d.

BGC	BGC type	Similar BGC	Notes
1	Other	None	–
2	RiPP (lantipeptide)	None	Single A-gene
3	Mixed NRPS- Type 1 PKS	None	transAT type PKS module
4a	RiPP (lantipeptide)	None	Two A-genes
4b	NRPS	None	Two siderophore like ORFs
4c	Type 3 PKS	None	Phloroglucinol synthase like
4d	Possible RiPP	None	Two radical SAM enzyme encoding genes
5	NRPS	None	Probable pentapeptide
6a	NRPS	Telomycin	Telomycin-like BGC
6b	Terpene	2-methyl isoborneol	2-methyl Isoborneol BGC
6c	Aminoglycoside	None	–
6d	NRPS	None	Probable decapeptide
7	NRPS	None	Probable pentapeptide
8	Mixed Type 1 PKS-NRPS	None	–
9	Mixed RiPP (lantipeptide) – Type 1 PKS	Abyssomicin	Single A-gene. RiPP genes directly adjacent to PKS genes.
10	Terpene	None	–
11	Ectoioine	Ectoioine	Ectoioine BGC
12	Mixed Type 1 PKS-NRPS-RiPP and deoxyhexose	None	RiPP, NRPS, PKS and deoxyhexose biosynthesis & transfer genes co-located. SapB BGC
13	RiPP	SapB	SapB BGC
14	Melanin	None	–
15	Siderophore	Desferrioxamine	Desferrioxamine B BGC
16	RiPP	None	–
17a	NRPS	Griseobactin	Griseobactin-like BGC
17b	NRPS	Lasparyomycin	Probable tridecapeptide
18	Mixed NRPS-FAS/PKS	None	–

(continued on next page)

Table 2 (continued)

BGC	BGC type	Similar BGC	Notes
19	Terpene	None	Pentalene synthase like
20	Terpene	Albaflavenone	Albaflavenone BGC
21	Siderophore	None	Aerobactin like
22	Butyrolactone	None	–
23	Bacteriocin	None	–
24	NRPS	Calcium dependent antibiotic	Probable lipo-hexapeptide
25	Terpene	Geosmin	Geosmin
26	Type 2 PKS	Fasamycin/formicamycin	Formicamycin BGC
27	Type 1 PKS-NRPS	None	–
28	Terpene	Hopenone	Hopenone BGC
29a	RiPP (lantipeptide)	None	Single A-gene
29b	Mixed Type 3 PKS-FAS	None	–
29c	RiPP (lassopeptide)	None	Single precursor gene
30	Type 1 PKS	None	Polyene-like; probable octadecaketide
31	NRPS	A40926	Glycopeptide antibiotic like
32	NRPS	None	–
33	Type 1 PKS	Lasaloic acid	Ionophore like; probable undecaketide;
34a	NRPS	None	Single module; terminal thioester reductase domain
34b	Clavam (beta-lactam)	Clavams	Clavam-like
34c	Type 1 PKS	None	–

*coelicolor* (Challis, 2013), and there are at least two siderophores BGCs, including one for desferrioxamine B.

### Conflicts of interest

The authors declare no conflicts of interest.

### Acknowledgements

We thank Juan Pablo Gomez-Escribano for helpful advice. This work was supported by the following awards: MRC Milstein award G0801721 (to MIH), NERC research grants NE/J01074X/1 (to MIH) and NE/M015033/1 (to MIH/BW), a Norwich Research Park (NRP) Translational Award (to BW/MIH), and a BBSRC NPRONET (BB/L013754/1) Proof of Concept award (to MIH/BW). RD is funded by a Norwich Research Park BBSRC Doctoral Training Program Studentship

BB/M011216/1 and ZQ is funded by the BBSRC via Institute Strategic Programme Grant BB/J004561/1 to the John Innes Centre.

### Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at <https://doi.org/10.1016/j.jbiotec.2017.11.011>.

### References

- Aziz, R.K., Bartels, D., Best, A.A., DeJongh, M., Disz, T., Edwards, R.A., Formsma, K., Gerdes, S., Glass, E.M., Kubal, M., Meyer, F., Olsen, G.J., Olson, R., Osterman, A.L., Overbeek, R.A., McNeil, L.K., Paarmann, D., Paczian, T., Parrello, B., Pusch, G.D., Reich, C., Stevens, R., Vassieva, O., Vonstein, V., Wilke, A., Zagnitko, O., 2008. The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9, 75. <http://dx.doi.org/10.1186/1471-2164-9-75>.
- Barrick, J.E., Colburn, G., Deatherage, D.E., Traverse, C.C., Strand, M.D., Borges, J.J., Knoester, D.B., Reba, A., Meyer, A.G., 2014. Identifying structural variation in haploid microbial genomes from short-read resequencing data using breseq. *BMC Genomics* 15, 1039. <http://dx.doi.org/10.1186/1471-2164-15-1039>.
- Blatrix, R., Djieto-Lordon, C., Mondolot, L., La Fisca, P., Voglmayr, H., McKey, D., 2012. Plant-ants use symbiotic fungi as a food Source: new insight into the nutritional ecology of ant-plant interactions. *Proc. Biol. Sci.* 279, 3940–3947. <http://dx.doi.org/10.1098/rspb.2012.1403>.
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., Madden, T.L., 2009. BLAST+: architecture and applications. *BMC Bioinform.* 10, 421–429. <http://dx.doi.org/10.1186/1471-2105-10-421>.
- Carver, T., Thomson, N., Bleasby, A., Berriman, M., Parkhill, J., 2009. DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics* 25, 119–120. <http://dx.doi.org/10.1093/bioinformatics/btn578>.
- Challis, G.L., 2013. Exploitation of the *Streptomyces coelicolor* A3(2) genome sequence for discovery of new natural products and biosynthetic pathways. *J. Ind. Microbiol. Biotechnol.* 41, 219–232. <http://dx.doi.org/10.1007/s10295-013-1383-2>.
- Gomez-Escribano, J.P., Castro, J.F., Razmilic, V., Chandra, G., Andrews, B., Asenjo, J.A., Bibb, M.J., 2015. The *Streptomyces leeuwenhoekii* genome: de novo sequencing and assembly in single contigs of the chromosome, circular plasmid pSLE1 and linear plasmid pSLE2. *BMC Genomics* 16, 485. <http://dx.doi.org/10.1186/s12864-015-1652-8>.
- Huang, C.H., Chen, C.Y., Tsai, H.H., Chen, C., Lin, Y.S., Chen, C.W., 2003. Linear plasmid SLP2 of *Streptomyces lividans* is a composite replicon. *Mol. Microbiol.* 47, 1563–1576. <http://dx.doi.org/10.1046/j.1365-2958.2003.03403.x>.
- Hyatt, D., Chen, G.L., Locascio, P.F., Land, M.L., Larimer, F.W., Hauser, L.J., 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinf.* 11, 119. <http://dx.doi.org/10.1186/1471-2105-11-119>.
- Kieser, T., Bibb, M.J., Buttner, M.J., Chater, K.F., Hopwood, D.A., 2000. *Practical Streptomyces Genetics*.
- Palmer, T.M., Stanton, M.L., Young, T.P., Goheen, J.R., Pringle, R.M., Karban, R., 2008. Breakdown of an ant-plant mutualism follows the loss of large herbivores from an African savanna. *Science* 319, 192–195. <http://dx.doi.org/10.1126/science.1151579>.
- Qin, Z., Munnoch, J.T., Devine, R., Holmes, N.A., Seipke, R.F., Wilkinson, K.A., Wilkinson, B., Hutchings, M.I., 2017. Formicamycins, antibacterial polyketides produced by *Streptomyces formicae* isolated from African *Tetraponera* plant-ants. *Chem. Sci.* 8, 3218–3227. <http://dx.doi.org/10.1039/C6SC04265A>.
- Seipke, R.F., Barke, J., Heavens, D., Yu, D.W., Hutchings, M.I., 2013. Analysis of the bacterial communities associated with two ant-plant symbioses. *Microbiol. Open* 2, 276–283. <http://dx.doi.org/10.1002/mb03.73>.