



Complete Genome Sequence of the Cryptophycin-Producing Cyanobacterium *Nostoc* sp. Strain ATCC 53789

Anna Tippelt,^a Tobias Busche,^b Christian Rückert,^c  Markus Nett^a

^aDepartment of Biochemical and Chemical Engineering, TU Dortmund University, Dortmund, Germany

^bIT Biotech GmbH Bielefeld, Bielefeld University, Bielefeld, Germany

^cTechnology Platform Genomics, Center for Biotechnology (CeBITec), Bielefeld University, Bielefeld, Germany

ABSTRACT *Nostoc* sp. strain ATCC 53789 is a producer of cryptophycins, which are promising anticancer agents. Here, we report the completely sequenced 8.7-Mb genome of *Nostoc* sp. strain ATCC 53789. The sequence provides insights into the metabolic network of this cyanobacterial strain and illuminates its potential for the biosynthesis of secondary metabolites.

The cyanobacterium *Nostoc* sp. strain ATCC 53789 produces two classes of bioactive secondary metabolites, i.e., the cryptophycins (1, 2) and the nostocyclopeptides (3). The cryptophycins were found to be potent anticancer agents (4, 5), which led to their clinical testing (6). Although the biosynthesis genes for cryptophycins and nostocyclopeptides were previously identified (7, 8), their integration into the metabolic network of the producing strain remained unclear.

To complement the existing information, the genome of *Nostoc* sp. strain ATCC 53789 was sequenced and assembled. Unless otherwise specified, default parameters were used for all software. The required DNA was isolated by phenol-chloroform extraction from a culture grown in BG-13 medium (1) under diurnal illumination for 1 month directly after receipt of the strain. The genome was reconstructed from short- and long-read DNA data sets obtained by Illumina and Nanopore sequencing. Library preparation involved a TruSeq DNA PCR-free high-throughput library prep kit (Illumina) and the SQK-LSK109 ligation sequencing kit (Oxford Nanopore Technologies [ONT]). Illumina sequencing was performed using a MiSeq reagent kit v3 (600 cycle) in a 2 × 300-nucleotide (nt) run. For Nanopore sequencing, a GridION platform with an R9.4.1 flow cell was used. Base calling and demultiplexing were performed using Guppy v3.1.5. Illumina data were assembled with Newbler v2.8 (9) (options: -large, -siom 16, -m, -consed). Nanopore data were processed with Canu v1.8 (10) (parameters: genomeSize = 6m, rawErrorRate = 0.3, correctedErrorRate = 0.1). Canu contigs were polished with Racon v1.3.3 (11) (parameters: -c 6, -m 8, -x -6, -g -8, -w 500), followed by medaka v0.11.0 (12) (parameters: -b 100, -m r941_min_high_g303) and Pilon v1.22 (13). minimap2 v2.17 (parameters: -ax sr, -secondary = no), BWA-MEM v2 (14) (parameters: -O1, -E1), and Bowtie 2 v2.3.2 (15) (parameters: -X 750, -no-unal) were used for mapping. Unicycler v0.4.6 (16) was used for hybrid assembly of the Illumina data and the contigs from the polished Canu assembly.

The assemblies were combined manually in Consed v27.0 (17). First, the chromosome was reoriented based on the *dnaA* gene. Overlapping ends from the polished Canu assembly were trimmed by assembly in Consed. Ambiguities in these regions as well as all other repeat regions were corrected based on the contigs produced by the Newbler assembly. Finally, all differences between the contigs of the three assemblies as well as low-quality regions marked in Consed were resolved by manual curation using IGV v2.4.14 (18) for visualization of the ONT data. This approach led to the

Citation Tippelt A, Busche T, Rückert C, Nett M. 2020. Complete genome sequence of the cryptophycin-producing cyanobacterium *Nostoc* sp. strain ATCC 53789. *Microbiol Resour Announc* 9:e00040-20. <https://doi.org/10.1128/MRA.00040-20>.

Editor J. Cameron Thrash, University of Southern California

Copyright © 2020 Tippelt et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

Address correspondence to Markus Nett, markus.nett@tu-dortmund.de.

Received 28 January 2020

Accepted 18 March 2020

Published 2 April 2020

TABLE 1 Genomic features of *Nostoc* sp. strain ATCC 53789

Replicon	Length (bp)	Topology	G+C content (%)	Coverage (×) for:		No. of biosynthetic loci ^a
				Nanopore data	Illumina data	
chr	7,340,101	Circular	41.4	298	39	18
pNsp_a	337,072	Circular	41.6	397	54	1
pNsp_b	325,114	Circular	40.9	330	43	1
pNsp_c	219,529	Circular	41.0	314	41	1
pNsp_d	65,222	Linear	41.8	314	62	1
pNsp_e	57,504	Circular	42.5	418	48	0
pNsp_f	56,077	Linear	41.7	528	98	0
pNsp_g	54,032	Circular	42.3	503	70	0
pNsp_h	49,561	Circular	39.6	361	61	0
pNsp_i	40,105	Circular	41.2	663	103	0
pNsp_j	38,437	Circular	41.4	450	94	0
pNsp_k	36,221	Circular	40.7	341	74	0
pNsp_l	34,754	Circular	42.0	349	64	0

^a According to antiSMASH.

identification of 13 replicons with a total size of 8,653,729 bp, including 1 circular chromosome, 10 circular plasmids, and 2 linear plasmids (Table 1). The plasmid topology was assessed by Canu and Unicycler and subsequently verified by inspection of the assemblies in Consed using the Illumina reads and the included to/fm/pr information provided by Newbler. The reads were mapped back onto the assembled contigs/replicons and checked in IGV for potential misassemblies. Genome annotation with Prokka v1.11 (19) resulted in the assignment of 7,408 genes, 7,300 protein-coding sequences, 88 tRNAs, 12 rRNAs, and 8 noncoding RNAs (ncRNAs).

An antiSMASH v5.0.0 (20) analysis revealed a distinctive secondary metabolome comprising 22 loci. Interestingly, the cryptophycin locus (7) resides on plasmid pNsp_c, flanked by transposase genes. Moreover, the analysis indicated that the strain is capable of anabaenopeptin biosynthesis (21).

Data availability. The annotated nucleotide sequences of the chromosome and the 12 plasmids of *Nostoc* sp. strain ATCC 53789 have been deposited at GenBank under the accession numbers [CP046703](#), [CP046704](#), [CP046705](#), [CP046706](#), [CP046707](#), [CP046708](#), [CP046709](#), [CP046710](#), [CP046711](#), [CP046712](#), [CP046713](#), [CP046714](#), and [CP046715](#). The raw data are available in the SRA under the accession numbers [SRR10969384](#) and [SRR10969385](#).

ACKNOWLEDGMENT

Financial support from the Bundesministerium für Bildung und Forschung (BMBF) is gratefully acknowledged (grant number 031B0356A).

REFERENCES

- Schwartz RE, Hirsch CF, Sesin DF, Flor JE, Chartrain M, Fromtling RE, Harris GH, Salvatore MJ, Liesch JM, Yudin K. 1990. Pharmaceuticals from cultured algae. *J Ind Microbiol* 5:113–123. <https://doi.org/10.1007/BF01573860>.
- Trimurtulu G, Ohtani I, Patterson GML, Moore RE, Corbett TH, Valeriote FA, Demchik L. 1994. Total structures of cryptophycins, potent antitumor depsipeptides from the blue-green alga *Nostoc* sp. strain GSV 224t. *J Am Chem Soc* 116:4729–4737. <https://doi.org/10.1021/ja00090a020>.
- Golakoti T, Yoshida WY, Chaganty S, Moore RE. 2001. Isolation and structure determination of nostocyclopeptides A1 and A2 from the terrestrial cyanobacterium *Nostoc* sp. ATCC53789. *J Nat Prod* 64:54–59. <https://doi.org/10.1021/np000316k>.
- Smith CD, Zhang X, Mooberry SL, Patterson GM, Moore RE. 1994. Cryptophycin: a new antimicrotubule agent active against drug-resistant cells. *Cancer Res* 54:3779–3784.
- Kerksiek K, Mejillano MR, Schwartz RE, Georg GI, Himes RH. 1995. Interaction of cryptophycin 1 with tubulin and microtubules. *FEBS Lett* 377:59–61. [https://doi.org/10.1016/0014-5793\(95\)01271-0](https://doi.org/10.1016/0014-5793(95)01271-0).
- Sessa C, Weigang-Köhler K, Pagani O, Greim G, Mora O, De Pas T, Burgess M, Weimer I, Johnson R. 2002. Phase I and pharmacological studies of the cryptophycin analogue LY355703 administered on a single intermittent or weekly schedule. *Eur J Cancer* 38:2388–2396. [https://doi.org/10.1016/s0959-8049\(02\)00489-6](https://doi.org/10.1016/s0959-8049(02)00489-6).
- Magarvey NA, Beck ZQ, Golakoti T, Ding Y, Huber U, Hemscheidt TK, Abelson D, Moore RE, Sherman DH. 2006. Biosynthetic characterization and chemoenzymatic assembly of the cryptophycins. Potent anticancer agents from *Nostoc* cyanobionts. *ACS Chem Biol* 1:766–779. <https://doi.org/10.1021/cb6004307>.
- Becker JE, Moore RE, Moore BS. 2004. Cloning, sequencing, and biochemical characterization of the nostocyclopeptide biosynthetic gene cluster: molecular basis for imine macrocyclization. *Gene* 325:35–42. <https://doi.org/10.1016/j.gene.2003.09.034>.
- Miller JR, Koren S, Sutton G. 2010. Assembly algorithms for next-generation sequencing data. *Genomics* 95:315–327. <https://doi.org/10.1016/j.ygeno.2010.03.001>.
- Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2017. Canu: scalable and accurate long-read assembly via adaptive k-mer

- weighting and repeat separation. *Genome Res* 27:722–736. <https://doi.org/10.1101/gr.215087.116>.
11. Vaser R, Sović I, Nagarajan N, Šikić M. 2017. Fast and accurate de novo genome assembly from long uncorrected reads. *Genome Res* 27:737–746. <https://doi.org/10.1101/gr.214270.116>.
 12. Oxford Nanopore Technologies. 2018. medaka. GitHub. <https://github.com/nanoporetech/medaka>.
 13. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, Earl AM. 2014. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One* 9:e112963. <https://doi.org/10.1371/journal.pone.0112963>.
 14. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv 1303.3997 [q-bio.GN]. <https://arxiv.org/abs/1303.3997>.
 15. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>.
 16. Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>.
 17. Gordon D, Abajian C, Green P. 1998. Consed: a graphical tool for sequence finishing. *Genome Res* 8:195–202. <https://doi.org/10.1101/gr.8.3.195>.
 18. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. 2011. Integrative genomics viewer. *Nat Biotechnol* 29:24–26. <https://doi.org/10.1038/nbt.1754>.
 19. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>.
 20. Blin K, Shaw S, Steinke K, Villebro R, Ziemert N, Lee SY, Medema MH, Weber T. 2019. antiSMASH 5.0: updates to the secondary metabolite genome mining pipeline. *Nucleic Acids Res* 47:W81–W87. <https://doi.org/10.1093/nar/gkz310>.
 21. Shishido TK, Jokela J, Fewer DP, Wahlsten M, Fiore MF, Sivonen K. 2017. Simultaneous production of anabaenopeptins and namalides by the cyanobacterium *Nostoc* sp. CENA543. *ACS Chem Biol* 12:2746–2755. <https://doi.org/10.1021/acschembio.7b00570>.