



# Genome Sequence of *Streptomyces* sp. Strain HB-N217, Isolated from the Marine Sponge *Forcepia* sp.

René K. M. Xavier,<sup>a</sup> Dongbo Xu,<sup>a</sup> Peter J. McCarthy,<sup>a</sup> Shengming Yang,<sup>b,c,d</sup>  Guojun Wang<sup>a</sup>

<sup>a</sup>Harbor Branch Oceanographic Institute, Florida Atlantic University, Fort Pierce, Florida, USA

<sup>b</sup>Department of Plant Pathology, North Dakota State University, Fargo, North Dakota, USA

<sup>c</sup>Department of Plant Sciences, North Dakota State University, Fargo, North Dakota, USA

<sup>d</sup>USDA-ARS Cereals Research Unit, Edward T. Schafer Agriculture Research Center, Fargo, North Dakota, USA

**ABSTRACT** The genome sequence of the *Forcepia* sponge-derived bacterium *Streptomyces* sp. strain HB-N217 was determined, with approximately 8.25 Mbp and a G+C content of 72.1%. Thirty biosynthetic gene clusters that bear the capability to produce secondary metabolites were predicted. The results will aid marine natural product chemistry and sponge-microbe association studies.

Actinomycetes, filamentous Gram-positive bacteria, are a rich source of secondary metabolites. Nearly half of the antibiotics in current use are produced by a single bacterial genus, *Streptomyces* (1). Since the determination of the first *Streptomyces* genome sequence and the realization of the abundance of cryptic gene clusters that encode enzymes for producing secondary metabolites (2), over 5,000 actinomycetal genomes have been determined; the cryptic gene clusters encoded therein have provided unprecedented opportunities for drug discovery (3).

The bacterial strain HB-N217 was cultivated using mucin agar plates (4) grown at 25°C for 1 month, from a sample (8-VIII-99-2-001) of the *Forcepia* sp. sponge which was collected at a depth of 70.5 m in the Gulf of Mexico, 103 miles west of Naples, FL, USA. In this study, HB-N217 was selected based on its preliminary identification as an actinomycete and its potential for production of secondary metabolites. The strain was grown in liquid soybean-peptone-yeast extract (SPY) medium for 72 h, and cells were collected and used for genomic DNA extraction using the cetyltrimethylammonium bromide (CTAB) method as described previously (5). Next, the 16S rRNA gene of HB-N217 was amplified from genomic DNA using the primers Ecoli9 and Loop27rc (6); BLAST analyses (7, 8) indicated that the HB-N217 16S gene was highly homologous with those of the genus *Streptomyces*, showing a 100% homology with a recently registered genome sequence, *Streptomyces* sp. strain NA03103 (GenBank accession number CP054920.1), suggesting that HB-N217 is a streptomycete.

Whole-genome sequencing was carried out at Genewiz using the Illumina MiSeq platform with 2 × 250-bp paired-end reads; the sequencing library was prepared by Genewiz according to the standard Illumina PCR-based library preparation kit. The assembled and annotated genome sequence was generated using a variety of quality-control and assembly methods using the Department of Energy Systems Biology Knowledgebase (KBase; <https://narrative.kbase.us/narrative/60713>) (9). The raw sequencing data were quality filtered with the JGI RQCFilter pipeline (BBTools v38.22) (10), followed by assembly with SPAdes v3.13.0 (11). QUAST v5.0.2 was used with the rna-finding parameter to generate assembly statistics and to predict rRNA genes (12). The completeness and contamination of the genome sequence was estimated using CheckM v1.0.18 via the lineage-specific workflow (13). Read alignment was performed using Bowtie 2 v2.4.2 in default mode (14). Taxonomic annotation of contigs was generated using the

**Citation** Xavier RKM, Xu D, McCarthy PJ, Yang S, Wang G. 2021. Genome sequence of *Streptomyces* sp. strain HB-N217, isolated from the marine sponge *Forcepia* sp. *Microbiol Resour Announc* 10:e01410-20. <https://doi.org/10.1128/MRA.01410-20>.

**Editor** Frank J. Stewart, Georgia Institute of Technology

**Copyright** © 2021 Xavier 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 Shengming Yang, [shengming.yang@usda.gov](mailto:shengming.yang@usda.gov), or Guojun Wang, [guojunwang@fau.edu](mailto:guojunwang@fau.edu).

**Received** 30 December 2020

**Accepted** 30 January 2021

**Published** 25 February 2021

Genome Taxonomy Database (GTDB-Tk) v1.0.2 with a minimum alignment of 10% (15). Genome annotation was done by NCBI's PGAP (Prokaryotic Genome Annotation Pipeline) (16, 17). The prophage was identified using VirSorter v1.0.2 and vConTACT2 v0.9.19 (18, 19). Default parameters were used for all software unless noted.

The genome assembly contained 331 contigs; the total contig length was 8,252,984 bp with a median G+C content of 72.1%, an  $N_{50}$  value of 41,623 bp, and a longest contig length of 153,966 bp. The genome sequence is classified as high quality, having 100% completeness and  $\leq 5\%$  contamination according to a recently published standard by the Genomic Standards Consortium (20); 96.87% of reads realigned to the assembly. A total of 7,474 genes, with 7,111 that encode proteins, were predicted in the genome, plus 1 noncoding CRISPR array, 4 noncoding CRISPRs, 3 noncoding CRISPR spacers, and 66 noncoding RNAs; one set of complete rRNA genes (5S, 16S, and 23S) and 68 tRNAs were found. Interestingly, a category 5 prophage was identified; however, it does not align with the viral genomes in the Prokaryotic Viral RefSeq v201 database (with ICTV and NCBI taxonomy).

In order to predict the secondary metabolic capability of HB-N217, antiSMASH v5.2.0 (21) was run to detect biosynthetic gene clusters (BCGs); 30 predictive BCGs were found in the HB-N217 genome sequence for the biosynthesis of diverse secondary metabolites, such as polyketides (e.g., pluramycin-type antimicrobials), nonribosomal peptides, terpenes, lanthipeptides, and so on, suggesting HB-N217 as a potential rich producer of marine natural products.

**Data availability.** The whole-genome assembly was deposited at NCBI under the accession number [JADWMQ000000000](https://www.ncbi.nlm.nih.gov/nuclseq/JADWMQ000000000). The version provided in the paper is the first version, [JADWMQ000000000.1](https://www.ncbi.nlm.nih.gov/nuclseq/JADWMQ000000000.1). The raw sequencing data have been deposited under the accession number [SRR13264572](https://www.ncbi.nlm.nih.gov/nuclseq/SRR13264572). A partial 16S rRNA gene sequence was deposited under the accession number [MT393585](https://www.ncbi.nlm.nih.gov/nuclseq/MT393585).

## ACKNOWLEDGMENTS

This work was supported by the Florida Atlantic University Harbor Branch Oceanographic Institute Foundation Faculty Start-up Package and National Institutes of Health grant R21CA209189 to G.W. and by U.S. Department of Agriculture–Agriculture Research Service (USDA-ARS) Current Research Information System (CRIS) project 3060-21000-038-00D to S.Y.

Shotgun sequencing and initial genome assembly were conducted by Genewiz, LLC.

## REFERENCES

- Barka EA, Vatsa P, Sanchez L, Gaveau-Vaillant N, Jacquard C, Klenk H-P, Clément C, Ouhdouch Y, van Wezel GP. 2016. Taxonomy, physiology, and natural products of Actinobacteria. *Microbiol Mol Biol Rev* 80:1–43. <https://doi.org/10.1128/MMBR.00019-15>.
- Bentley SD, Chater KF, Cerdeño-Tárraga A-M, Challis GL, Thomson NR, James KD, Harris DE, Quail MA, Kieser H, Harper D, Bateman A, Brown S, Chandra G, Chen CW, Collins M, Cronin A, Fraser A, Goble A, Hidalgo J, Hornsby T, Howarth S, Huang C-H, Kieser T, Larke L, Murphy L, Oliver K, O'Neil S, Rabinowitsch E, Rajandream M-A, Rutherford K, Rutter S, Seeger K, Saunders D, Sharp S, Squares R, Squares S, Taylor K, Warren T, Wietzorrek A, Woodward J, Barrell BG, Parkhill J, Hopwood DA. 2002. Complete genome sequence of the model actinomycete *Streptomyces coelicolor* A3(2). *Nature* 417:141–147. <https://doi.org/10.1038/417141a>.
- Reddy TBK, Thomas AD, Stamatis D, Bertsch J, Isbandi M, Jansson J, Mallajosyula J, Pagani I, Lobos EA, Kyrpides NC. 2015. The Genomes OnLine Database (GOLD) v.5: a metadata management system based on a four level (meta)genome project classification. *Nucleic Acids Res* 43: D1099–D1106. <https://doi.org/10.1093/nar/gku950>.
- Olson JB, McCarthy PJ. 2005. Associated bacterial communities of two deep-water sponges. *Aquat Microb Ecol* 39:47–55. <https://doi.org/10.3354/ame039047>.
- Kieser T, Bibb MJ, Chater KF, Buttner MJ, Hopwood DA. 2000. Practical *Streptomyces* genetics: a laboratory manual. John Innes Foundation, Norwich, United Kingdom.
- Sfanos K, Harmody D, Dang P, Ledger A, Pomponi S, McCarthy P, Lopez J. 2005. A molecular systematic survey of cultured microbial associates of deep-water marine invertebrates. *Syst Appl Microbiol* 28:242–264. <https://doi.org/10.1016/j.syapm.2004.12.002>.
- Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol* 215:403–410. [https://doi.org/10.1016/S0022-2836\(05\)80360-2](https://doi.org/10.1016/S0022-2836(05)80360-2).
- Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: architecture and applications. *BMC Bioinformatics* 10:421. <https://doi.org/10.1186/1471-2105-10-421>.
- Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, Sneddon MW, Henderson ML, Riehl WJ, Murphy-Olson D, Chan SY, Kamimura RT, Kumari S, Drake MM, Brettin TS, Glass EM, Chivian D, Gunter D, Weston DJ, Allen BH, Baumohl J, Best AA, Bowen B, Brenner SE, Bun CC, Chandonia J-M, Chia J-M, Colasanti R, Conrad N, Davis JJ, Davison BH, DeJongh M, Devoid S, Dietrich E, Dubchak I, Edirisinghe JN, Fang G, Faria JP, Frybarger PM, Gerlach W, Gerstein M, Greiner A, Gurtowski J, Haun HL, He F, Jain R, et al. 2018. KBase: the United States Department of Energy Systems Biology Knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>.
- Bushnell B, Rood J, Singer E. 2017. BBMerge—accurate paired shotgun read merging via overlap. *PLoS One* 12:e0185056. <https://doi.org/10.1371/journal.pone.0185056>.
- Nurk S, Bankevich A, Antipov D, Gurevich A, Korobeynikov A, Lapidus A, Prjibelsky A, Pyshkin A, Sirotkin A, Sirotkin Y, et al. 2013. Assembling genomes and mini-metagenomes from highly chimeric reads, p 158–170.

- In Deng M, Jiang R, Sun F, Zhang X (ed), Research in computational molecular biology. RECOMB 2013. Lecture notes in computer science, vol 7821. Springer, Berlin, Germany.
- Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. *Bioinformatics* 29:1072–1075. <https://doi.org/10.1093/bioinformatics/btt086>.
  - Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>.
  - Langmead B, Wilks C, Antonescu V, Charles R. 2019. Scaling read aligners to hundreds of threads on general-purpose processors. *Bioinformatics* 35:421–432. <https://doi.org/10.1093/bioinformatics/bty648>.
  - Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>.
  - Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, Lomsadze A, Pruitt KD, Borodovsky M, Ostell J. 2016. NCBI Prokaryotic Genome Annotation Pipeline. *Nucleic Acids Res* 44:6614–6624. <https://doi.org/10.1093/nar/gkw569>.
  - Haft DH, DiCuccio M, Badretdin A, Brover V, Chetvernin V, O'Neill K, Li W, Chitsaz F, Derbyshire MK, Gonzales NR, Gwadz M, Lu F, Marchler GH, Song JS, Thanki N, Yamashita RA, Zheng C, Thibaud-Nissen F, Geer LY, Marchler-Bauer A, Pruitt KD. 2018. RefSeq: an update on prokaryotic genome annotation and curation. *Nucleic Acids Res* 46:D851–D860. <https://doi.org/10.1093/nar/gkx1068>.
  - Roux S, Enault F, Hurwitz BL, Sullivan MB. 2015. VirSorter: mining viral signal from microbial genomic data. *PeerJ* 3:e985. <https://doi.org/10.7717/peerj.985>.
  - Bolduc B, Jang HB, Doulier G, You Z-Q, Roux S, Sullivan MB. 2017. vConTACT: an iVirus tool to classify double-stranded DNA viruses that infect Archaea and Bacteria. *PeerJ* 5:e3243. <https://doi.org/10.7717/peerj.3243>.
  - Bowers RM, Kyrpides NC, Stepanauskas R, Harmon-Smith M, Doud D, Reddy TBK, Schulz F, Jarett J, Rivers AR, Eloe-Fadrosh EA, Tringe SG, Ivanova NN, Copeland A, Clum A, Becraft ED, Malmstrom RR, Birren B, Podar M, Bork P, Weinstock GM, Garrity GM, Dodsworth JA, Yooseph S, Sutton G, Glöckner FO, Gilbert JA, Nelson WC, Hallam SJ, Jungbluth SP, Etema TJG, Tighe S, Konstantinidis KT, Liu W-T, Baker BJ, Rattei T, Eisen JA, Hedlund B, McMahon KD, Fierer N, Knight R, Finn R, Cochrane G, Karsch-Mizrachi I, Tyson GW, Rinke C, Lapidus A, Meyer F, Yilmaz P, Parks DH, Murat Eren A, The Genome Standards Consortium, et al. 2017. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. *Nat Biotechnol* 35:725–731. <https://doi.org/10.1038/nbt.3893>.
  - 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>.