Title

Genome Sequences of Actinobacteria from Extreme Environments in Colombia

Journal

Microbiology Resource Announcements, 7(22)

ISSN

2576-098X

Authors

Cantillo, Angela

Shapiro, Nicole

Woyke, Tanja

et al.

Publication Date

2018-12-06

DOI

10.1128/mra.01384-18

Peer reviewed

AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

Genome Sequences of Actinobacteria from Extreme Environments in Colombia

Angela Cantillo,[a] Nicole Shapiro,[b] Tanja Woyke,[b] Nikos C. Kyrpides,[b] Sandra Baena,[c] María Mercedes Zambrano[a]

[a]Molecular Genetics, Corporación Corpogen–Research Center, Bogotá, Colombia

[b]Department of Energy, Joint Genome Institute, Walnut Creek, California, USA

[c]Unidad de Saneamiento y Biotecnología Ambiental, Departamento de Biología, Pontificia Universidad Javeriana, Bogotá, Colombia

ABSTRACT We sequenced six actinobacterial genomes isolated from a salt mine and from soil in a high-mountain Páramo ecosystem. The strains belonged to the genera *Streptomyces*, *Nesterenkonia*, and *Isoptericola* and were sequenced due to their antimicrobial and cytotoxic activities.

Actinobacteria produce a broad range of natural products, most of which have been obtained from actinomycetes belonging to the genus *Streptomyces* (1). Given the limitation of phenotypic screening and the rediscovery of previously isolated compounds, genome mining of microorganisms isolated from unique environments is attractive for the identification of novel biosynthetic gene clusters (BGCs) (2). Genomic analysis has shown that actinobacteria can harbor multiple BGCs that are not expressed in laboratory assays but may still have biological activities and industrial potential (3). The strains sequenced here were isolated from two locations of extreme environmental conditions in Colombia to expand the actinobacterial collection for biological activity assays (3). Samples were collected from a salt mine (Zipaquirá, 5°01′06.18″N, 74°0′13.63″W) (4), located at 2,656 m above sea level (masl), pH 6.6, and temperature of 17.3°C, and from an Andean Páramo ecosystem (5) soil in the Nevados Natural National Park (04°50′55.7″N, 75°21′51.3″W), located at 4,141 masl, with pH 5.6 and temperature of 7.2°C. We isolated three soil *Streptomyces* strains on Difco Actinomycete medium and three *Nesterenkonia* and *Isoptericola* halotolerant strains from the salt mine by growth on tryptic soy agar (Merck) supplemented with 8% NaCl, as described previously (4). The strains were genome sequenced to characterize and compare their metabolic potential, based on their previously characterized capacity to inhibit growth of Gram-positive and Gram-negative bacteria, eukaryotic microbes, and their cytotoxic activity against cell lines 4T1 (mouse mammary tumor) and MCF-7 (human mammary adenocarcinoma) (4) (Table 1).

DNAs were extracted as reported previously (6) by centrifuging an overnight culture, lysing cells with 200-μm glass beads in the presence of cetyltrimethylammonium bromide (CTAB)-NaCl in a FastPrep (2 × 30 s) (MP Biomedical, Santa Ana, CA), followed by phenol-chloroform extraction and DNA precipitation with isopropanol. Draft genomes were obtained at the DOE Joint Genome Institute by constructing 300-bp-insert shotgun libraries that were barcoded and sequenced as pools of three libraries on the Illumina MiSeq platform (2 × 150-bp paired ends). All raw sequence data were filtered using BBDuk version 35.83 (7), which eliminates known Illumina artifacts and PhiX. Reads with more than one "N" or with quality scores averaging less than 8 (before trimming) or reads shorter than 51 bp (after trimming) were discarded. The remaining reads were mapped using BBMap version 35.83 (7) to human, cat, and dog references masked using BBMask version 35.83 (7). Assembly was carried out as follows: (i) artifact-filtered Illumina reads were assembled using Velvet (version 1.2.07) (8), (ii) 1 to

Received 9 October 2018 **Accepted** 6 November 2018 **Published** 6 December 2018

Citation Cantillo A, Shapiro N, Woyke T, Kyrpides NC, Baena S, Zambrano MM. 2018. Genome sequences of actinobacteria from extreme environments in Colombia. Microbiol Resour Announc 7:e01384-18. https://doi.org/10.1128/MRA.01384-18.

Editor Frank J. Stewart, Georgia Institute of Technology

Address correspondence to María Mercedes Zambrano, mzambrano@corpogen.org.

TABLE 1 Strain and genome characteristics[a]

Strain (GenBank accession no.)	Biological activity			No. of reads	No. of contigs/N_{50} (bp)	No. of proteins	No. of RNAs	No. of genes	GC content (%)	Length (Mb)	Coverage (×)	No. of BGCs
	Bacterial	Fungal	Cytotoxic									
Streptomyces avidinii CG 885 (QLTM00000000)	+ (G+, G−)	+	−	10,452,432	45/588,328	6,783	90	6,873	71.08	7.65	204	55
Streptomyces microflavus CG 893 (OAOR00000000)	+ (G+, G−)	−	−	9,842,212	45/623,933	6,551	93	6,644	71.12	7.37	199	51
Streptomyces sp. CG 926 (QGGZ00000000)	G+	+	ND	12,394,618	46/470,107	7,552	108	7,660	71.83	8.51	176	72
Isoptericola sp. strain CG 20/1183 (PVTW00000000)	−	−	+4T1	11,274,090	26/263,847	3,501	68	3,569	73.10	3.87	378	17
Isoptericola halotolerans CG 23/1184 (PVTX00000000)	−	−	+4T1, MCF7	11,400,520	22/280,589	3,486	69	3,555	73.11	3.85	382	18
Nesterenkonia sandarakina CG35/1185 (PVTY00000000)	−	−	+4T1, MCF7	11,167,940	58/180,192	2,983	62	3,045	67.46	3.22	454	13

[a] Growth inhibition was absent (−) or present (+) when assayed against Gram-positive (G+) or Gram-negative (G−) bacteria or fungi. Cytotoxic activity according to reference 4. ND, not determined.

3-kbp simulated paired-end reads were created from the Velvet contigs using Wgsim (version 0.3.0) (9), and (iii) Illumina reads were assembled with simulated read pairs using AllPaths-LG (version r46652) (10). Parameters for the assembly steps were (i) Velvet (velveth, 63 –shortPaired; velvetg, –very clean yes –exportFiltered yes –min contig lgth 500 –scaffolding no –cov cutoff 10), (ii) wgsim (–e 0 –1 100 –2 100 –r 0 –R 0 –X 0), and (iii) AllPaths-LG (PrepareAllpathsInputs, PHRED 64 = 0 PLOIDY = 1 FRAG COVERAGE = 125 JUMP COVERAGE = 25 LONG JUMP COV = 50; RunAllpathsLG, THREADS = 8 RUN=std shredpairs TARGETS=standard VAPI WARN ONLY=True OVERWRITE=True). Annotation was performed using the DOE-JGI annotation pipeline version 4.10.5 (11), BGCs were identified using antiSMASH version 2.0.2 (12) and the IMG-ABC database (13), using default settings in both cases, and taxonomic identities were assigned based on 16S rRNA gene analysis.

The number of reads per genome ranged between 9,842,212 and 12,394,618. The N_{50} values ranged from 180,192 to 623,933 bp (Table 1). All genomes had a high GC content (67 to 71%) but differed in size, with the soil *Streptomyces* strains having genomes more than twice the size of those of the three halotolerant strains. The number of genes varied from 3,045 in *Nesterenkonia sandarakina* to 7,660 for *Streptomyces* sp. strain CG 926. The three *Streptomyces* genomes also had more BGCs (Table 1) predicted to encode various specialized metabolites. The differences in genome length and coding features for the two *Isoptericola* strains and three *Streptomyces* strains also indicate distinct metabolic potential in isolates of the same genus.

Data availability. These draft genome sequences have been deposited in GenBank under the accession numbers QLTM00000000, OAOR00000000, QGGZ00000000, PVTW00000000, PVTX00000000, and PVTY00000000. The data are available in the NCBI Sequence Read Archive under accession numbers SRX2947591, SRX3047884, SRX3047887, SRX2947595, SRX2947597, and SRX2947593.

ACKNOWLEDGMENTS

This work was partially funded by Colciencias (grants 120356934285 and 275898511). Strain collection and sequencing were performed under permits from Park Authority (DTNO-N-20/2007) and the Ministry of the Environment (numbers 15 and 104). The work conducted by the U.S. Department of Energy Joint Genome Institute, a DOE Office of Science User Facility, is supported by the Office of Science of the U.S. Department of Energy under contract number DE-AC02-05CH11231.

We thank Laura Bohórquez and Katherine Carrillo for help with isolation of the strains.

A.C. performed biological activity assays and genomic preparations. N.S., T.W., and N.C.K. participated in the genome sequencing, assembly, and annotation. S.B. and M.M.Z. conceived of the study, participated in its design and coordination, and helped draft the manuscript. All authors read and approved the final manuscript.

REFERENCES

1. Bérdy J. 2012. Thoughts and facts about antibiotics: where we are now and where we are heading. J Antibiot (Tokyo) 65:385–395. https://doi.org/10.1038/ja.2012.27.
2. Tracanna V, de Jong A, Medema MH, Kuipers OP. 2017. Mining prokaryotes for antimicrobial compounds: from diversity to function. FEMS Microbiol Rev 41:417–429. https://doi.org/10.1093/femsre/fux014.
3. Genilloud O. 2014. The re-emerging role of microbial natural products in antibiotic discovery. Antonie Van Leeuwenhoek 106:173–188. https://doi.org/10.1007/s10482-014-0204-6.
4. Díaz-Cárdenas C, Cantillo A, Rojas LY, Sandoval T, Fiorentino S, Robles J, Ramos FA, Zambrano MM, Baena S. 2017. Microbial diversity of saline environments: searching for cytotoxic activities. AMB Express 7:223. https://doi.org/10.1186/s13568-017-0527-6.
5. Ruiz-Pérez CA, Restrepo S, Zambrano MM. 2016. Microbial and functional diversity within the phyllosphere of Espeletia sp. in an Andean high mountain ecosystem. Appl Environ Microbiol 82:1807–1817. https://doi.org/10.1128/AEM.02781-15.
6. Cayol JL, Ollivier B, Patel BK, Ravot G, Magot M, Ageron E, Grimont PA, Garcia JL. 1995. Description of Thermoanaerobacter brockii subsp. lactiethylicus subsp. nov., isolated from a deep subsurface French oil well, a proposal to reclassify Thermoanaerobacter finnii as Thermoanaerobacter brockii subsp. finnii comb. nov., and an emended description of Thermoanaerobacter brockii. Int J Syst Bacteriol 45:783–789. https://doi.org/10.1099/00207713-45-4-783.
7. Bushnell B. 2017. BBTools software package. https://sourceforge.net/projects/bbmap/.
8. Zerbino DR, Birney E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18:821–829. https://doi.org/10.1101/gr.074492.107.
9. Li H. 2011. Wgsim. https://github.com/lh3/wgsim.
10. Gnerre S, Maccallum I, Przybylski D, Ribeiro FJ, Burton JN, Walker BJ, Sharpe T, Hall G, Shea TP, Sykes S, Berlin AM, Aird D, Costello M, Daza R, Williams L, Nicol R, Gnirke A, Nusbaum C, Lander ES, Jaffe DB. 2011. High-quality draft assemblies of mammalian genomes from massively

parallel sequence data. Proc Natl Acad Sci U S A 108:1513–1518. https://doi.org/10.1073/pnas.1017351108.

11. Huntemann M, Ivanova NN, Mavromatis K, Tripp HJ, Paez-Espino D, Palaniappan K, Szeto E, Pillay M, Chen I-MA, Pati A, Nielsen T, Markowitz VM, Kyrpides NC. 2015. The standard operating procedure of the DOE-JGI Microbial Genome Annotation Pipeline (MGAP v.4). Stand Genomic Sci 10:86. https://doi.org/10.1186/s40793-015-0077-y.
12. Weber T, Blin K, Duddela S, Krug D, Kim HU, Bruccoleri R, Lee SY, Fischbach MA, Müller R, Wohlleben W, Breitling R, Takano E, Medema MH. 2015. antiSMASH 3.0—a comprehensive resource for the genome mining of biosynthetic gene clusters. Nucleic Acids Res 43:W237–W243. https://doi.org/10.1093/nar/gkv437.
13. Hadjithomas M, Chen IMA, Chu K, Huang J, Ratner A, Palaniappan K, Andersen E, Markowitz V, Kyrpides NC, Ivanova NN. 2017. IMG-ABC: new features for bacterial secondary metabolism analysis and targeted biosynthetic gene cluster discovery in thousands of microbial genomes. Nucleic Acids Res 45:D560–D565. https://doi.org/10.1093/nar/gkw1103.