CrossMark
click for updates

Whole-Genome Sequences of Five Oligotrophic Bacteria Isolated from Deep within Lechuguilla Cave, New Mexico

Huan You Gan,[a] Han Ming Gan,[a] Alexander Mario Tarasco,[b] Nurfatini Idayu Busairi,[b] Hazel A. Barton,[c] André O. Hudson,[b] Michael A. Savka[b]

School of Science, Monash University Malaysia, Jalan Lagoon Selatan, Bandar Sunway, Petaling Jaya, Selangor, Malaysia[a]; The Thomas H. Gosnell School of Life Sciences, Rochester Institute of Technology, Rochester, New York, USA[b]; Department of Biology, University of Akron, Akron, Ohio, USA[c]

Here, we report the whole-genome sequences and annotation of five oligotrophic bacteria from two sites within the Lechuguilla Cave in the Carlsbad Caverns National Park, NM. Three of the five genomes contain an acyl-homoserine lactone signal synthase ortholog (*luxI*) that is involved in cell-to-cell communication via quorum sensing.

Address correspondence to Michael A. Savka, massbi@rit.edu.

Bacterial diversity from extreme oligotrophic environments, such as caves, is of interest given the unique substrate environment. As such, we embarked on a project to isolate and identify bacteria from two sites in the Lechuguilla Cave, NM (1). The genomes of bacterial species from caves as deep as Lechuguilla Cave (>400 m) have not yet been sequenced, with only three cave bacterial genomes sequenced to date: *Beutenbergia cavernae* from Guangxi, China, *Pseudomonas fluorescens* from the Guiana Shield, South America, and *Gloeobacter kilaueensis* from a biofilm in Kīlauea Caldera, Hawaii (2–4). The bacterial genomes presented here were isolated from cave surfaces: strain LC238 was collected from an iron- and manganese-rich corrosion residue at the contact between the limestone bedrock and a back reef sandstone (the Yates Formation) at −200 m, while LC5, LC81, LC85, and LC363 were collected from a deep and remote location within the limestone Capitan Formation at −347 m. Each site is extremely nutrient limited, with organic carbon present at between 0.1 and 0.7 mg/g of sediment and with a C-to-N ratio approaching 300:1 (1). The bacteria were initially identified by nucleotide sequence analysis of the variable regions 3 and 4 of the 16S rRNA gene (1), with LC5 and LC363 identified as members of the genera *Devosia* and *Sphingopyxis*, respectively (Table 1).

The sequencing libraries were prepared from extracted DNA using the Nextera XT kit (Illumina, San Diego, CA). Each sample was tagged with unique bar codes, according to the manufacturer's protocol. The final libraries were normalized based on the 2100 Bioanalyzer readings (Agilent Technologies) and pooled for sequencing on the MiSeq sequencing system (Illumina) at the Monash University Malaysia Genomics Facility to generate FASTQ files. Raw FASTQ reads for each library were corrected for errors and *de novo* assembled into contigs with the SPAdes Genome Assembler (version 2.5.1) (5). Scaffolds were then generated from the assembled contigs using SSPACE (version 2.0) (6). The gaps in the resulting scaffolds were then closed using GapFiller (version 1.11) (7). The annotation for each genome was performed using the Prokka (version 1.8) annotation pipeline (8), which comprises Aragorn (version 1.2.36), Prodigal (version 2.60), and RNAmmer (version 1.2), which predicted tRNAs, open reading frames (ORFs), and rRNAs, respectively (9–11). The predicted 16S rRNA from RNAmmer was queried using BLASTn against the NCBI nt database. The genus of each sample sequence was determined by manually observing the distance tree of the result to check if the query sequence falls within a certain cluster of organisms (12). InterProScan 5 was used to provide additional annotation to the predicted protein sequences (13). A summary of the key properties for each genome is shown in Table 1. Phylogenetic analysis of

TABLE 1 Sequencing and annotation results of the five cave bacteria isolated from two sites within the Lechuguilla Cave in the Carlsbad Caverns National Park, NM[a]

Strain	Source (depth in m)	BioProject no.	Biosample no.	Accession no.	Organism	Genome coverage (×)	Genome size (bp)	No. of contigs	No. of ORFs	No. of tRNAs	No. of rRNAs
LC5	LCEAE (−347)	PRJNA248423	SAMN02798393	JNNO00000000	*Devosia* sp.	65	4,202,991	47	4,117	48	6
LC81	LCEAE (−347)	PRJNA248597	SAMN02799681	JNFD00000000	*Sphingopyxis* sp.	104	4,397,290	48	4,111	44	3
LC85	LCEAE (−347)	PRJNA248600	SAMN02799685	JPKG00000000	*Bosea* sp.	47	6,564,029	74	6,193	51	3
LC238	LCAE1 (−200)	PRJNA248601	SAMN02799686	JNNN00000000	*Massilia* sp.	29	5,799,774	83	5,102	65	7
LC363	LCEAE (−347)	PRJNA248602	SAMN02799687	JNFC00000000	*Sphingopyxis* sp.	28	4,210,757	73	3,908	48	3

[a] Sites where samples were obtained in the Lechuguilla Cave, Carlsbad Caverns National Park, New Mexico (refer to Fig. 1, profile map of the Lechuguilla Cave system, in reference 1).

the full-length 16S rRNA gene reclassified strain LC238 as a *Massilia* species. Further, genome analysis demonstrated that strains LC81, LC363, and LC238 contain a *luxI* homolog, implicating that cell-to-cell communication is potentially mediated by acyl-homoserine lactones in cave environments.

Nucleotide sequence accession numbers. The nucleotide sequences have been deposited at DDBJ/EMBL/GenBank under the accession numbers provided in Table 1.

REFERENCES

1. **Bhullar K, Waglechner N, Pawlowski A, Koteva K, Banks ED, Johnston MD, Barton HA, Wright GD.** 2012. Antibiotic resistance is prevalent in an isolated cave microbiome. PLoS One **7**:e34953. http://dx.doi.org/10.1371/journal.pone.0034953.
2. **Land M et al.** 2009. Complete genome sequence of *Beutenbergia cavernae* type strain (HKI 0122ᵗ). Genomics Standards **1**:21–28. http://dx.doi.org/10.4056/sigs.1162.
3. **Barton MD, Petronio M, Giarrizzo JG, Bowling BV, Barton HA.** 2013. The genome of *Pseudomonas fluorescens* strain R124 demonstrates phenotypical adaptation to the mineral environment. J. Bacteriol. **195:** 4793–4803. http://dx.doi.org/10.1128/JB.00825-13.
4. **Saw JHW, Schatz M, Brown MV, Kunkel DD, Foster JS, Shick H, Christensen S, Hou S, Wan X, Donachie SP.** 2013. Cultivation and complete genome sequencing of *Gloeobacter kilaueensis* sp. nov., from the lava cave in Kilauea Caldera, Hawaii. PLoS One **8**:e76376. http://dx.doi.org/10.1371/journal.pone.0076376.
5. **Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA.** 2012. SPAdes: a new genome assembly algorithm and its applications to single cell sequencing. J. Comput. Biol. **19**:455–477. http://dx.doi.org/10.1089/cmb.2012.0021.
6. **Boetzer M, Henkel CV, Jansen HJ, Butler D, Pirovano W.** 2011. Scaffolding pre-assembled contigs using SSPACE. Bioinformatics **27**: 578–579. http://dx.doi.org/10.1093/bioinformatics/btq683.
7. **Boetzer M, Pirovano W.** 2012. Toward almost closed genomes with GapFiller. Genome Biol. **13**:2012–2013. http://dx.doi.org/10.1186/gb-2012-13-6-r56.
8. **Seeman T.** 2014. Prokka: rapid prokayotic genome annotation. Bioinformatics **30**:2068–2069. http://dx.doi.org/10.1093/bioinformatics/btu153.
9. **Laslett D, Canback B.** 2004. Aragorn, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. Nucleic Acids Res. **32**:11–16. http://dx.doi.org/10.1093/nar/gkh152.
10. **Hyatt D, Chen GL, Locascio PF, Land ML, Larimer FW, Hauser LJ.** 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinformatics **11**:119. http://dx.doi.org/10.1186/1471-2105-11-119.
11. **Lagesen K, Hallin P, Rødland EA, Staerfeldt HH, Rognes T, Ussery DW.** 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. **35**:3100–3108. http://dx.doi.org/10.1093/nar/gkm160.
12. **Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ.** 1990. Basic local alignment search tool. J. Mol. Biol. **215**:403–410. http://dx.doi.org/10.1016/S0022-2836(05)80360-2.
13. **Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A, Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R, Hunter S.** 2014. InterProScan 5: genome-scale protein function classification. Bioinformatics. http://dx.doi.org/10.1093/bioinformatics/btu031.