AMERICAN SOCIETY FOR MICROBIOLOGY | Microbiology Resource Announcements

GENOME SEQUENCES

"*Candidatus* Colwellia aromaticivorans" sp. nov., "*Candidatus* Halocyntiibacter alkanivorans" sp. nov., and "*Candidatus* Ulvibacter alkanivorans" sp. nov. Genome Sequences

Mariana E. Campeão,[a,b] Jean Swings,[a,b,c] Bruno Sergio Silva,[a,b] Koko Otsuki,[a,b] Fabiano L. Thompson,[a,b] Cristiane C. Thompson[a]

[a]Laboratory of Microbiology, Instituto de Biologia, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[b]Center of Technology-CT2, SAGE-COPPE, Federal Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil
[c]Ghent University, Ghent, Belgium

ABSTRACT Unplanned oil spills during offshore production are a serious problem for the industry and the marine environment. Here, we present the genome sequence analysis of three novel hydrocarbon-degrading bacteria, namely, "*Candidatus* Colwellia aromaticivorans" sp. nov., "*Candidatus* Halocyntiibacter alkanivorans" sp. nov., and "*Candidatus* Ulvibacter alkanivorans" sp. nov.

Metagenomic data obtained from microcosm experiments with oil incubation in seawater from the Foz do Amazonas and Barreirinhas basins indicated the biodegradation potential of deep-sea microbial communities (1). Here, we present the metagenome assembled genome (MAG) sequences of three novel hydrocarbon-degrading bacteria obtained from these previous microcosm experiments, as detailed in our previous work (1). First, read trimming was obtained with prinseq v0.20.4 (-trim_qual_right/trim_qual_left 20, -trim_qual_type min, -trim_qual_rule lt, -trim_qual_window 5, -trim_qual_step 1, -min_len 35, -min_qual_mean 20, -trim_left/-trim_right 10, -out_bad null, -derep 1, and -custom_params\"CTGTCTCTTATACACATCT 1;CTGTCTCTTATACACATCTGACGCTGCCGACGA 1;CTGTCTCTTATACACATCTCCGAGCCCACGAGAC 1\) (2). Subsequently, a binning approach was performed through cross-assembly using eight metagenomes from Foz do Amazonas and eight metagenomes from Barreirinhas with SPAdes v3.6.2 (3); using k-mers of 21, 33, 55, 77, 99, and 127; mapping reads against assembled scaffolds with Bowtie 2 v2.2.5 (4) (–very-fast, -a, –no-unal, –no-discordant, and –no-mixed); converting to bam format with SAMtools v1.7 (5); coverage estimating with jgi_summarize_bam_contig_depths (–outputDepth); and binning with MetaBAT v0.25.4 (–specific, -v, -m 2500, and –minSamples 5) (6). RefineM v0.0.23 (scaffold_stats, outliers, and filter_bins commands) (7) was used to remove contamination, and CheckM v1.0.11 (lineage_wf workflow) (8) was used to assess quality stats. Genomic taxonomy was performed as described previously (9–12). These three novel MAGs showed less than 98.7% 16S rRNA sequence identity with the closest species. The three novel lineages had <95% amino acid identity (AAI) and <70% DNA-DNA hybridization (DDH) toward the closest phylogenetic neighbors, suggesting that the recovered genomes are new species, namely, "*Candidatus* Colwellia aromaticivorans" sp. nov., "*Candidatus* Halocyntiibacter alkanivorans" sp. nov., and "*Candidatus* Ulvibacter alkanivorans" sp. nov. (Table 1). "*Candidatus* Ulvibacter alkanivorans" sp. nov. and "*Candidatus* Halocyntiibacter alkanivorans" sp. nov. possess an alkane monooxygenase, the specific enzyme for alkane degradation, that was lacking in "*Candidatus* Colwellia aromaticivorans" sp. nov. "*Candidatus* Colwellia aromaticivorans" has a phenol 2-monooxygenase gene involved in toluene biodegradation and harbors two mono-

Citation Campeão ME, Swings J, Silva BS, Otsuki K, Thompson FL, Thompson CC. 2019. "*Candidatus* Colwellia aromaticivorans" sp. nov., "*Candidatus* Halocyntiibacter alkanivorans" sp. nov., and "*Candidatus* Ulvibacter alkanivorans" sp. nov. genome sequences. Microbiol Resour Announc 8:e00086-19. https://doi.org/10.1128/MRA.00086-19.

Editor Catherine Putonti, Loyola University Chicago

Address correspondence to Fabiano L. Thompson, fabianothompson1@gmail.com, or Cristiane C. Thompson, thompsoncristiane@gmail.com.

Received 24 January 2019
Accepted 3 March 2019
Published 11 April 2019

TABLE 1 Genomic features of metagenome-associated genomes[a]

Genome feature	Data for strain:		
	"*Candidatus* Colwellia aromaticivorans" sp. nov.	"*Candidatus* Halocynthiibacter alkanivorans" sp. nov.	"*Candidatus* Ulvibacter alkanivorans" sp. nov.
Coverage (×)	97	19	19
Completeness (%)	97.4	98.7	93.1
Contamination (%)	1.8	2.15	0.2
Size (Mbp)	3.71	4.7	2.7
No. of contigs	101	171	117
N_{50} value (bp)	72,862	55,069	47,750
No. of ORFs[b]	3,278	4,540	2,625
16S rRNA similarity with closest relative (%)	*Colwellia rossensis* S51-W(gv)1^T (98.5)	*Halocynthiibacter arcticus* PAMC 20958^T (97.2)	*Ulvibacter antarcticus* IMCC3101^T (96.2)
GGD[c] with closest relative (%)	*Colwellia rossensis* S51-W(gv)1^T (ND)[d]	*Halocynthiibacter arcticus* PAMC 20958^T (19.4)	*Ulvibacter antarcticus DSM 23424* (17.8)
AAI with closest relative (%)	*Colwellia rossensis* S51-W(gv)1^T (ND)	*Halocynthiibacter arcticus* PAMC 20958^T (65.4)	*Ulvibacter antarcticus* DSM 23424 (71.8)

[a] 16S rRNA sequences recovered were retrieved with RNAmmer (14) and compared with BLAST (15) against the Silva database (16, 17). The closest were selected to align through ssu-align v0.1.1 (18, 19) and perform phylogeny through IQ-TREE (20–22). The 16S rRNA similarity was calculated with TaxonDC (23). AAI and DDH values were performed with CompareM (https://github.com/dparks1134/CompareM) and GGDC (http://ggdc.dsmz.de/), respectively. The genome of *Ulvibacter antarcticus* DSM 23424 was used for AAI and DDH analysis. *In silico*-predicted phenotype analysis reveals that "*Candidatus* Colwellia aromaticivorans" sp. nov. is positive for chitin hydrolysis, propionate utilization, and DL-lactate utilization. "*Candidatus* Halocynthiibacter alkanivorans" sp. nov. is positive for acid production from D-ribose, sucrose, and glycerol; acetate utilization; indole production; and nitrate reduction. "*Candidatus* Ulvibacter alkanivorans" sp. nov. is positive for DL-lactic acid utilization.
[b] ORFs, open reading frames.
[c] GGD, genome-to-genome distance.
[d] ND, not determined.

oxygenases and nine oxidoreductases. Oxygenases, a class of oxidoreductases, are key enzymes in hydrocarbon degradation (13).

***"Candidatus* Colwellia aromaticivorans" (ar.o.ma'ti.ca. N.L. adj. *aromaticivorans*, aromatic, referring to the property of utilizing aromatic compounds).** *In silico*-predicted phenotypes indicate that the novel species hydrolyzes chitin, utilizes propionate and DL-lactate, and does not utilize D-fructose, L-arabinose, glycerol, citrate, or L-malate. The G+C content is 38%.

***"Candidatus* Halocyntiibacter alkanivorans" (al.ka.ni'vo.rans. M.L. n. *alkanum* saturated aliphatic hydrocarbon; L. v. *vorare* to eat; L. adj. *alkanivorans* alkane-devouring).** *In silico*-predicted phenotypes indicate that the novel species produces acid from D-ribose, sucrose, and glycerol. Acetate is utilized, nitrate is reduced, and indole is produced. The G+C content is 59%.

***"Candidatus* Ulvibacter alkanivorans" (al.ka.ni'vo.rans. M.L. n. *alkanum* saturated aliphatic hydrocarbon; L. v. *vorare* to eat; L. adj. *alkanivorans* alkane-devouring).** *In silico*-predicted phenotypes indicate that the novel species utilizes DL-lactate and does not hydrolyze starch or carboxymethyl cellulose (CM-cellulose). It is negative for the utilization of D-mannitol, cellobiose, D-fructose, L-fucose, D-galactose, myo-inositol, D-sorbitol, sucrose, trehalose, and propionic acid. The G+C content is 40%.

Data availability. This whole-genome shotgun project has been deposited in GenBank under the BioProject accession no. PRJNA478776.

ACKNOWLEDGMENTS

We thank CNPQ, CAPES, and FAPERJ for the support.

REFERENCES

1. Campeão ME, Reis L, Leomil L, de Oliveira L, Otsuki K, Gardinali P, Pelz O, Valle R, Thompson FL, Thompson CC. 2017. The deep-sea microbial community from the Amazonian basin associated with oil degradation. Front Microbiol 8:1019. https://doi.org/10.3389/fmicb.2017.01019.
2. Schmieder R, Edwards R. 2011. Quality control and preprocessing of metagenomic datasets. Bioinformatics 27:863–864. https://doi.org/10.1093/bioinformatics/btr026.
3. 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, Pyshkin AV. 2012. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol 19:455–477. https://doi.org/10.1089/cmb.2012.0021.
4. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. Nat Methods 9:357–359. https://doi.org/10.1038/nmeth.1923.
5. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis

G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The Sequence Alignment/Map format and SAMtools. Bioinformatics 25:2078–2079. https://doi.org/10.1093/bioinformatics/btp352.
6. Kang DD, Froula J, Egan R, Wang Z. 2015. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ 3:e1165. https://doi.org/10.7717/peerj.1165.
7. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. Nat Microbiol 2:1533–1542. https://doi.org/10.1038/s41564-017-0012-7.
8. 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.
9. Amaral GRS, Dias GM, Wellington-Oguri M, Chimetto L, Campeão ME, Thompson FL, Thompson CC. 2014. Genotype to phenotype: identification of diagnostic vibrio phenotypes using whole genome sequences. Int J Syst Evol Microbiol 64:357–365. https://doi.org/10.1099/ijs.0.057927-0.
10. Amaral GRS, Campeão ME, Swings J, Thompson FL, Thompson CC. 2015. Finding diagnostic phenotypic features of *Photobacterium* in the genome sequences. Antonie Van Leeuwenhoek 107:1351–1358. https://doi.org/10.1007/s10482-015-0414-6.
11. Thompson CC, Amaral GR, Campeão M, Edwards RA, Polz MF, Dutilh BE, Ussery DW, Sawabe T, Swings J, Thompson FL. 2015. Microbial taxonomy in the post-genomic era: rebuilding from scratch? Arch Microbiol 197:359–370. https://doi.org/10.1007/s00203-014-1071-2.
12. De Vos P, Thompson F, Thompson C, Swings J. 2017. A flavor of prokaryotic taxonomy: systematics revisited, p 29–44. *In* Kurtböke I (ed), Microbial resources: from functional existence in nature to applications. Elsevier B.V., Amsterdam, Netherlands.
13. Vandecasteele JP, Jones T. 2008. Petroleum microbiology: concepts, environmental implications, industrial applications. Editions Technips, Paris, France.
14. Lagesen K, Hallin P, Rødland EA, Stærfeldt H-H, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res 35:3100–3108. https://doi.org/10.1093/nar/gkm160.
15. 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.
16. Pruesse E, Quast C, Knittel K, Fuchs BM, Ludwig W, Peplies J, Glöckner FO. 2007. SILVA: a comprehensive online resource for quality checked and aligned ribosomal RNA sequence data compatible with ARB. Nucleic Acids Res 35:7188–7196. https://doi.org/10.1093/nar/gkm864.
17. Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO. 2013. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res 41:D590–D596. https://doi.org/10.1093/nar/gks1219.
18. Nawrocki EP, Kolbe DL, Eddy SR. 2009. Infernal 1.0: inference of RNA alignments. Bioinformatics 25:1335–1337. https://doi.org/10.1093/bioinformatics/btp157.
19. Cannone JJ, Subramanian S, Schnare MN, Collett JR, D'Souza LM, Du Y, Feng B, Lin N, Madabusi LV, Müller KM, Pande N, Shang Z, Yu N, Gutell RR. 2002. The Comparative RNA Web (CRW) site: an online database of comparative sequence and structure information for ribosomal, intron, and other RNAs. BMC Bioinformatics 3:2. https://doi.org/10.1186/1471-2105-3-2.
20. Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. 2015. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum likelihood phylogenies. Mol Biol Evol 32:268–274. https://doi.org/10.1093/molbev/msu300.
21. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. 2018. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol 35:518–522. https://doi.org/10.1093/molbev/msx281.
22. Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14:587–589. https://doi.org/10.1038/nmeth.4285.
23. Tarlachkov SV, Starodumova IP. 2017. TaxonDC: calculating the similarity value of the 16S rRNA gene sequences of prokaryotes or ITS regions of fungi. J Bioinform Genom 3:1–4. https://doi.org/10.18454/jbg.2017.3.5.1.