Draft genome sequence of three hydrocarbon-degrading *Pseudomonadota* strains isolated from an abandoned century-old oil exploration well

Jefferson Camacho,[1,2] Esteve Mesén-Porras,[1] Diego Rojas-Gätjens,[1] Danilo Pérez-Pantoja,[3,4] Fernando Puente-Sánchez,[5] Max Chavarría[1,2,6]

AUTHOR AFFILIATIONS   See affiliation list on p. 3.

ABSTRACT   We present genome sequences of three *Pseudomonadota* strains isolated from an abandoned century-old oil exploration well. A *Pseudomonas* sp. genome showed a size of 5,378,420 bp, while *Acinetobacter* genomes sized 3,522,593 and 3,864,311 bp. Genomes included catabolic genes for benzoate, 4-hydroxybenzoate, salicylate, vanillate, indoleacetate, anthranilate, *n*-alkanes, 4-hydroxyphenylacetate, phenylacetate, among others.

Studies focused in hydrocarbon bioremediation have highlighted the potential of exploring microorganisms containing genes related to hydrocarbon-degrading enzymes (1, 2). Specifically, bacteria belonging to genera *Pseudomonas* (3) and *Acinetobacter* (4) have shown potential to degrade alkanes and aromatic hydrocarbons (5, 6). Here, we report the genomic sequencing of three strains retrieved from an abandoned old-century oil exploration well, located in the Cahuita National Park, Costa Rica, which is characterized by a continuous efflux of methane and the presence of a complex mixture of hydrocarbons (7). Sampling details and isolation methods are described by Rojas-Gätjens et al. (7). Such microorganisms were identified as one strain of *Pseudomonas* sp. termed C11 and two different *Acinetobacter* strains termed C4 and C10.

Isolates stored in glycerol 20% at –80°C were grown in lysogeny broth (8) (150 mL) and incubated overnight (180 rpm, 30°C). Genomic DNA extraction was performed using cetyltrimethylammonium bromide protocol (9). DNA was sequenced at Novogene company Ltd., Singapore. A paired-end sequencing library was prepared using the Illumina HiSeq Preparation Kit and loaded onto NovaSeq PE150 system. The raw data quality was evaluated using FastQC v0.11.9 (10), and low-quality reads and adapter sequences were filtered using Trimmomatic v0.36 (11). The remaining reads were assembled with SPADES v5.2.0 (12), followed by assembly polishing through Pilon v1.24 (13). Contigs smaller than 500 bp were removed with BBMap suite v37.36 (14). The quality assessment of each assembly was summarized using QUAST v5.2.0 (15). Genome annotation was determined using NCBI Prokaryotic Genome Annotation Pipeline (16). Default parameters were used for all software. Genome assembly statistics and annotation features are presented in Table 1.

Taxonomic analysis was performed using the tetra correlation search tool of JspeciesWS v4.0.2 (17). This analysis showed that C4 strain is related to *Acinetobacter johnsonii*, C10 is closely related to *Acinetobacter pittii,* and C11 only can be classified as *Pseudomonas* sp. Consistently, average nucleotide identity (18) showed that C4 and C10 share more than 96% and 99% of its genome with *A. johnsonii* and *A. pittii* strains,
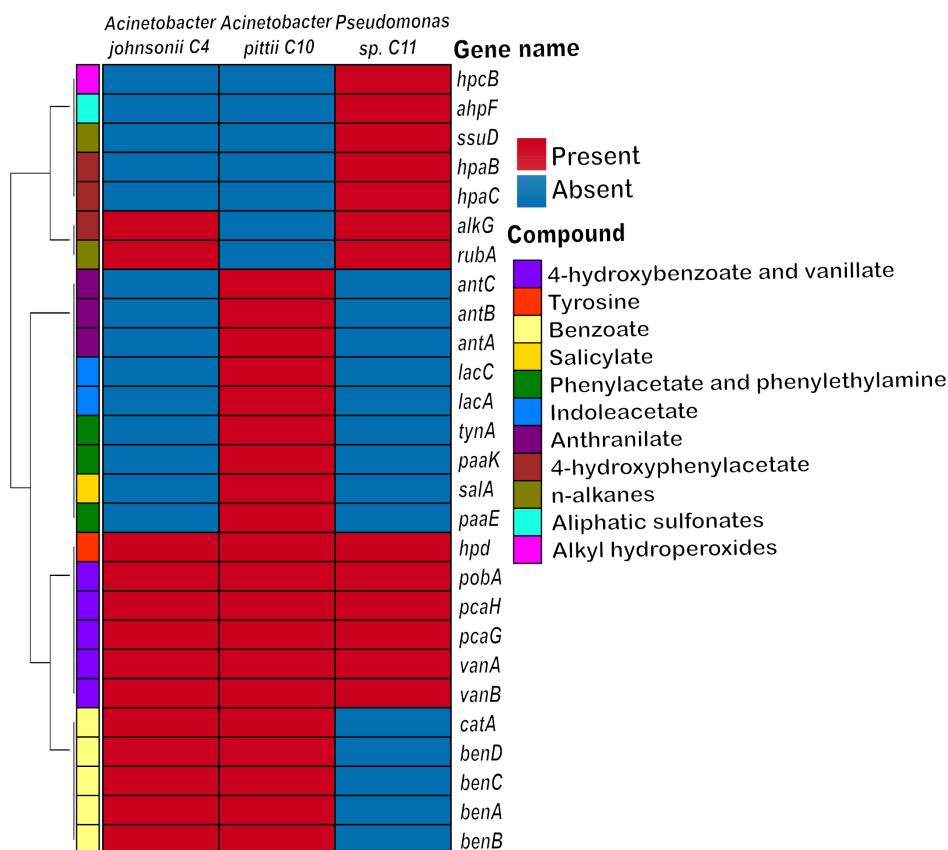
FIG 1 Genes related to hydrocarbons biodegradation identified in three bacterial strains isolated from an abandoned century-old oil exploration well in Cahuita National Park, Costa Rica. The color scale in the heatmap indicates the absence (light blue color) or presence (reddish color) of 27 different catabolic genes related to the biodegradation of specific aromatic or aliphatic hydrocarbons. The gene names are indicated on the right of the heatmap. The dendrogram on the left represents the clustering of the genes based on the Manhattan algorithm. Gene description of the protein product: *hpcB*, homoprotocatechuate 2,3-dioxygenase; *ahpF*, alkyl hydroperoxide reductase subunit F; *ssuD*, alkanesulfonate monooxygenase; *hpaB*, 4-hydroxyphenylacetate 3-monooxygenase, oxygenase component; *hpaC*, 4-hydroxyphenylacetate 3-monooxygenase, reductase component; *alkG*, rubredoxin, subunit 2; *rubA*, rubredoxin, subunit 1; *antC*, anthranilate 1,2-dioxygenase, reductase component; *antB*, anthranilate 1,2-dioxygenase, small subunit of terminal oxygenase component; *antA*, anthranilate 1,2-dioxygenase, large subunit of terminal oxygenase component; *lacC*, laccase domain-containing protein 1; *lacA*, galactoside acetyltransferase; *tynA*, 2-phenylethylamine oxidase; *paaK*, phenylacetate-coenzyme A ligase; *salA*, salicylate 1-hydroxylase; *paaE*, 1,2-phenylacetyl-CoA epoxidase, subunit E; *hpd*, 4-hydroxyphenylpyruvate dioxygenase; *pobA*, phenoxybenzoate dioxygenase, alpha subunit; *pcaH*, protocatechuate 3,4, dioxygenase, beta subunit; *pcaG*, protocatechuate 3,4, dioxygenase, alpha subunit; *vanA*, vanillate O-demethylase, oxygenase subunit; *vanB*, vanillate O-demethylase, oxidoreductase subunit; *catA*, catechol 1,2-dioxygenase; *benD*, 1,6-dihydroxycyclohexa-2,4-diene-1-carboxylate dehydrogenase; *benC*, benzoate 1,2-dioxygenase, electron transfer component; *benA*, benzoate 1,2-dioxygenase, alpha subunit; *benB*, benzoate 1,2-dioxygenase, beta subunit.

respectively. Nonetheless, there were no identity values above the species threshold (>95%) for C11, suggesting that such isolate could represent a new *Pseudomonas* species. To evaluate the metabolic capacities of each strain, we analyzed the proteome generated by genome annotation to identify enzymes related to hydrocarbons biodegradation by using the Hydrocarbon Aerobic Degradation Enzymes and Genes database (19) and the ProteinOrtho v6.3.0 tool (20). A matrix for plotting a heatmap was built using tidyverse (21), RColorBrewer v1.1–3 (22), colorRamps v2.3.1, and ComplexHeatmap v2.14.0 (23). The results are showed in Fig. 1, revealing the presence of genes related to aromatic compounds degradation, including benzoate, 4-hydroxybenzoate, vanillate, salicylate, anthranilate, indoleacetate, and phenylacetate as well medium- and long-chain alkanes.

TABLE 1 Annotation and genome assembly statistics of three hydrocarbon-degrading strains belonging to *Pseudomonas* and *Acinetobacter* genera isolated from an abandoned century-old oil exploration well

Isolate Feature	*Pseudomonas* sp. C11	*Acinetobacter johnsonii* C4	*Acinetobacter pittii* C10
BioProject accession no.	PRJNA997083	PRJNA997083	PRJNA997083
Assembly accession no.	JAWIIP000000000.1	JAVQMB000000000.1	JAWIIO000000000.1
No. of raw reads	1,175,0304	7,211,368	7,795,620
Genome size (bp)	5,378,420	3,522,593	3,864,311
No. of contigs	35	59	14
No. of predicted genes	4,943	2,009	1,812
Coverage ($1\times$) (bp)	51,813	78,898	72,299
G + C content (%)	62.85	41.44	38.72
N_{50} (bp)	312,100	113,976	2,118,397
No. of tRNA	60	63	61
No. of rRNA	3	3	3
No. of ncRNA	33	11	30

AUTHOR AFFILIATIONS

[1]Centro Nacional de Innovaciones Biotecnológicas (CENIBiot), CeNAT-CONARE, San José, Costa Rica

[2]Escuela de Química, Universidad de Costa Rica, San José, Costa Rica

[3]Instituto Universitario de Investigación y Desarrollo Tecnológico, Universidad Tecnológica Metropolitana, Santiago, Chile

[4]Center of Applied Ecology and Sustainability (CAPES), Santiago, Chile

[5]Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, Lennart Hjelms väg, Uppsala, Sweden

[6]Centro de Investigaciones en Productos Naturales (CIPRONA), Universidad de Costa Rica, San José, Costa Rica

AUTHOR ORCIDs

Max Chavarría http://orcid.org/0000-0001-5901-3576

AUTHOR CONTRIBUTIONS

Jefferson Camacho, Investigation, Methodology | Esteve Mesén-Porras, Data curation, Formal analysis, Methodology, Software, Validation, Writing – original draft | Diego Rojas-Gätjens, Conceptualization, Investigation, Methodology | Danilo Pérez-Pantoja, Conceptualization, Data curation, Supervision, Validation, Writing – review and editing | Fernando Puente-Sánchez, Data curation, Supervision, Validation, Writing – review and editing | Max Chavarría, Conceptualization, Funding acquisition, Project administration, Resources, Supervision, Writing – review and editing

DATA AVAILABILITY

The whole-genome sequencing shotgun for the three microorganisms was submitted to NCBI/GenBank under BioProject PRJNA997083. Raw reads were deposited with SRA numbers SRR25386955, SRR25386957, SRR25386958 and the assembled genomes with accessions JAWIIP000000000.1, JAWIIO000000000.1, JAVQMB000000000.1. The current version referred to this paper is identified as version 1.0.

REFERENCES

1. Liang C, Ye Q, Huang Y, Wang Y, Zhang Z, Wang H. 2022. Shifts of the new functional marker gene (*pahE*) of polycyclic aromatic hydrocarbons (PAHs) degrading bacterial population and its relationship with pahs biodegradation. J Hazard Mater 437:129305. https://doi.org/10.1016/j.jhazmat.2022.129305

2. Ravi A, Ravuri M, Krishnan R, Narenkumar J, Anu K, Alsalhi MS, Devanesan S, Kamala-Kannan S, Rajasekar A. 2022. Characterization of petroleum degrading bacteria and its optimization conditions on effective utilization of petroleum hydrocarbons. Microbiol Res 265:127184. https://doi.org/10.1016/j.micres.2022.127184

3. Wang X, Jin D, Zhou L, Wu L, Qi L, Li C, An W, Chen Y. 2014. Draft genome sequence of halotolerant polycyclic aromatic hydrocarbon-degrading *Pseudomonas bauzanensis* strain W13Z2. Genome Announc 2:e01049-14. https://doi.org/10.1128/genomeA.01049-14

4. Macaya CC, Méndez V, Durán RE, Aguila-Torres P, Salvà-Serra F, Jaén-Luchoro D, Moore ERB, Seeger M. 2019. Complete genome sequence of hydrocarbon-degrading halotolerant *Acinetobacter radioresistens* DD78,Isolated from the aconcagua river mouth in central Chile. Microbiol Resour Announc 8:e00601-19. https://doi.org/10.1128/MRA.00601-19

5. Freitas JF, Silva DFL, Silva BS, Castro JNF, Felipe M, Silva-Portela RCB, Minnicelli CF, Agnez-Lima LF. 2023. Genomic and phenotypic features of *Acinetobacter baumannii* isolated from oil reservoirs reveal a novel subspecies specialized in degrading hazardous hydrocarbons. Microbiol Res 273:127420. https://doi.org/10.1016/j.micres.2023.127420

6. Sui X, Li Y, Wang X, Yu L, Wang J, Ji H. 2023. Genomics and degradation law of *Acinetobacter junii* in response to petroleum pollution. Process Biochemistry 126:41–50. https://doi.org/10.1016/j.procbio.2022.12.021

7. Rojas-Gätjens D, Fuentes-Schweizer P, Rojas-Jiménez K, Pérez-Pantoja D, Avendaño R, Alpízar R, Coronado-Ruíz C, Chavarría M. 2022. Methylotrophs and hydrocarbon-degrading bacteria are key players in the microbial community of an abandoned century-old oil exploration well. Microb Ecol 83:83–99. https://doi.org/10.1007/s00248-021-01748-1

8. Bertani G. 2004. Lysogeny at mid-twentieth century: P1, P2, and other experimental systems. J Bacteriol 186:595–600. https://doi.org/10.1128/JB.186.3.595-600.2004

9. Anderson KL, Lebepe-Mazur S. 2003. Comparison of rapid methods for the extraction of bacterial DNA from colonic and caecal lumen contents of the pig. J Appl Microbiol 94:988–993. https://doi.org/10.1046/j.1365-2672.2003.01917.x

10. Andrews S. 2010. FastQC: a quality control tool for high throughput sequence data. Babraham Inst. Retrieved 23 Feb 2021. https://www.bioinformatics.babraham.ac.uk/projects/fastqc.

11. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for illumina sequence data. Bioinform 30:2114–2120. https://doi.org/10.1093/bioinformatics/btu170

12. 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. https://doi.org/10.1089/cmb.2012.0021

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. 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

15. Gurevich A, Saveliev V, Vyahhi N, Tesler G. 2013. QUAST: quality assessment tool for genome assemblies. Bioinform 29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086

16. 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

17. Richter M, Rosselló-Móra R, Oliver Glöckner F, Peplies J. 2016. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. Bioinform 32:929–931. https://doi.org/10.1093/bioinformatics/btv681

18. Ciufo S, Kannan S, Sharma S, Badretdin A, Clark K, Turner S, Brover S, Schoch CL, Kimchi A, DiCuccio M. 2018. Using average nucleotide identity to improve taxonomic assignments in prokaryotic genomes at the NCBI. Int J Syst Evol Microbiol 68:2386–2392. https://doi.org/10.1099/ijsem.0.002809

19. Rojas-Vargas J, Castelán-Sánchez HG, Pardo-López L. 2023. HADEG: a curated hydrocarbon aerobic degradation enzymes and genes database. Comput Biol Chem 107:107966. https://doi.org/10.1016/j.compbiolchem.2023.107966

20. Lechner M, Findeiss S, Steiner L, Marz M, Stadler PF, Prohaska SJ. 2011. Proteinortho: detection of (Co-)orthologs in large-scale analysis. BMC Bioinform 12:124. https://doi.org/10.1186/1471-2105-12-124

21. Couch S, Bray A, Ismay C, Chasnovski E, Baumer B, Çetinkaya-Rundel M. 2021. Infer: an R package for tidyverse-friendly statistical inference. JOSS 6:3661. https://doi.org/10.21105/joss.03661

22. Neuwirth E. 2014. Rcolorbrewer: colorbrewer palettes. R package 1:1–2.

23. Gu Z, Eils R, Schlesner M. 2016. Complex heatmaps reveal patterns and correlations in multidimensional genomic data. Bioinform 32:2847–2849. https://doi.org/10.1093/bioinformatics/btw313