

Draft genome of a multi-drug, metal and biocide-resistant strain of *Klebsiella variicola* isolated from the Yaque del Sur River in the Dominican Republic

David Tavares Martins,^{1,2,3} Oscar Victor Cardenas Alegria,^{1,2} Thais Azevedo Cardoso,^{1,2} Antônio Márcio Gomes Martins-Júnior,⁴ Luiz Orlando Maroto-Martin,⁵ Edian Franco Franklin de Los Santos,^{5,6,7} Rommel Thiago Jucá Ramos¹

AUTHOR AFFILIATIONS See affiliation list on p. 3.

ABSTRACT *Klebsiella variicola* strain YSP4_INTEC was isolated from a Dominican Republic River and had its genome sequenced (size: 5,611,899 bp and GC content: 57.19%). The genomic analysis identified various genes involved in plant colonization mechanisms, plus a wide array of multi-drug, biocide, and multi-metal resistance genes.

KEYWORDS *Klebsiella variicola*, resistance, multi-drug, Caribbean

Klebsiella variicola is a highly adaptable bacterial species found in diverse environments and maintains complex mutualistic or parasitic relationships with plants, fungi, or animals (1). It exhibits plant growth-promoting genes and can remove heavy metals and phenolic compounds from the environment (2–5). However, some strains are highly virulent and resistant to antibiotics, causing severe infections in humans (6–8). Due to limitations in automated microbiology equipment, differentiating *Klebsiella pneumoniae* from *K. variicola* remains challenging (2). This difficulty highlights the need for genomic studies, especially in places like the Dominican Republic, where data on related infections are scarce.

River water samples were collected from Tamayo, Batoruco (18°23.962'N, 71°11.165'W). The water was serially diluted (10⁻¹ to 10⁻⁴) and plated on MacConkey Agar, incubated at 37°C for 24 h. Colonies were isolated on the same medium and cultured on CHROMagar Orientation. The COL-APSE and SuperCARB culture media were employed to identify colistin-resistant and carbapenem-resistant colonies, respectively. The DNA was extracted from a single colony with the DNeasy Blood and Tissue (QIAGEN) Kit and quantified with a Qubit fluorometer. Sequencing was performed by the NovoGene company and conducted in the Illumina NovaSeq6000 platform with a microbial whole genome library (2 × 150 bp paired-end) without customizations. Quality control was conducted using AdapterRemoval (3.0.0) (9), KmerStream (1.1) (10), and FastQC (0.12.1) (11). Genome assembly utilized Edena (3.131028) (12), SPAdes (4.2.0) (13), and Unicycler (0.5.1) (14), with the best assembly selected based on QUAST (5.3.0) (15) results. Annotation was performed using Prokka (1.14.6) (16), and MOB-recon (3.1.9) (17) was used for plasmid annotation and extraction. These steps were performed via the Assembly Hi-Seq pipeline (18). Taxonomic classification was followed using GTDB-Tk v2 (v1.6.0) (19). Annotation was performed using PATRIC (<https://www.bv-brc.org/>) (20), and AMR identification was performed with AMR++ (v2.0.2) (21) and CARD (v6.0.6) (22). VFAnalyzer (<https://www.mgc.ac.cn/VFs/main.htm>) (23) was used for virulence gene identification. CheckM (v1.0.18) (24) was used for completeness evaluation. Phylogenetic trees were reconstructed using a maximum likelihood criterion in RaxML-NG, comparing (25) the isolate to 19 *Klebsiella* species genomes from NCBI. All the tools were used with default parameters.

Editor Vanja Klepac-Ceraj, Wellesley College, Wellesley, Massachusetts, USA

Address correspondence to Rommel Thiago Jucá Ramos, rommelthiago@gmail.com.

The authors declare no conflict of interest.

Received 29 June 2025

Accepted 7 September 2025

Published 24 October 2025

Copyright © 2025 Martins 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/).

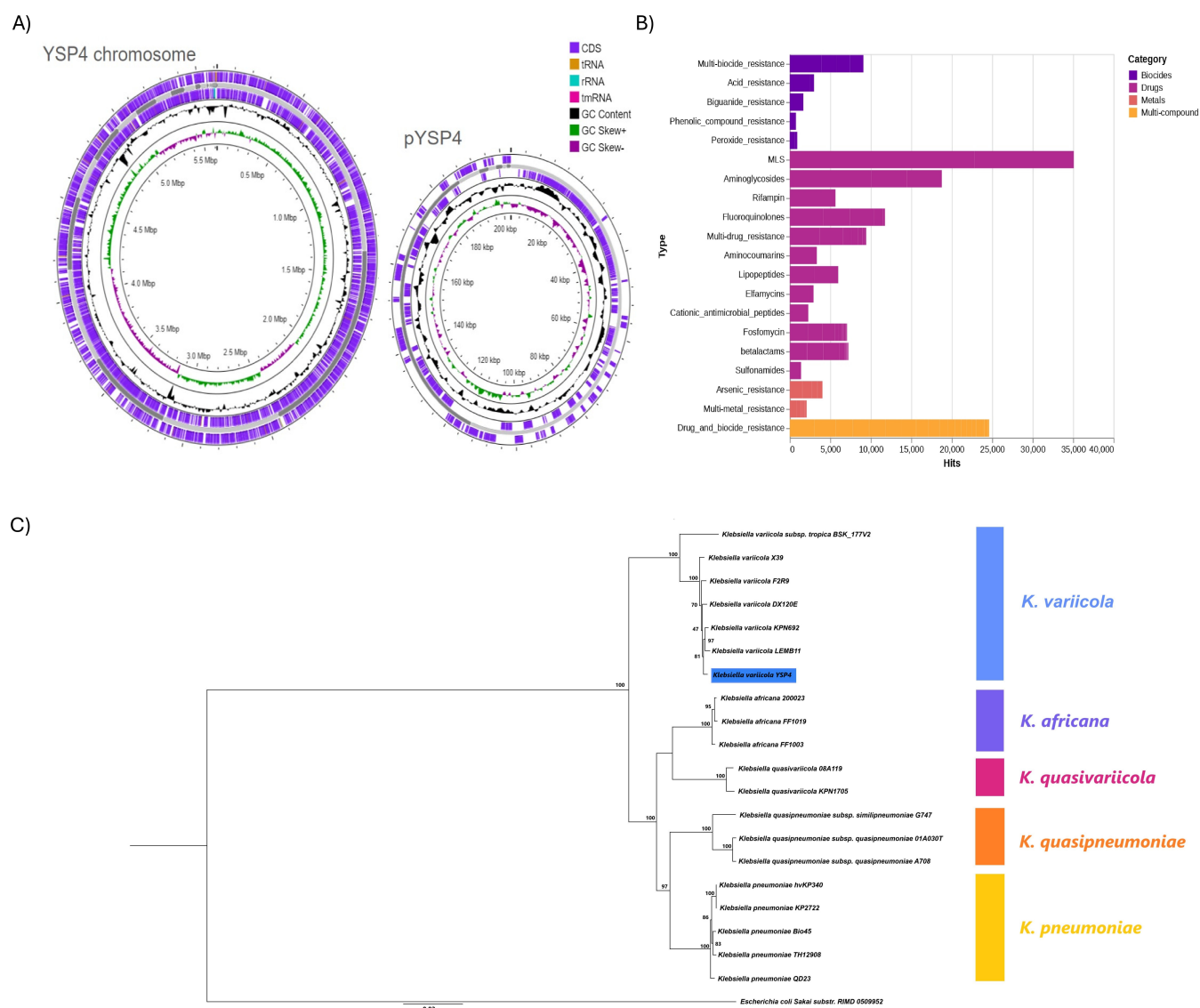


FIG 1 (A) Circular map of chromosome and plasmid from *K. variicola* strain YSP4_INTEC's annotated. From outside to the center: genes on forward strand, genes on reverse strand, GC content, and GC skew. Circular map was generated by Proksee. (B) Reads of YSP4_INTEC's sequencing distribution over four different categories of bacterial resistance genes in the MEGARES database analyzed by AMR plus plus. (C) Phylogenetic tree of the *Klebsiella* group genome sequences showing the position of *K. variicola* strain YSP4_INTEC. *Escherichia coli* Sakai substr. RIMD 0509952 is used as an outgroup. The values located in front or behind the tree nodes represent the bootstrap support value.

The analysis revealed a 5,611,899 bp chromosome assembled in 60 contigs (N50 = 715,876 bp, L50 = 3, and GC content = 57.19%) with 228× coverage, 99.61% completeness, and 9,690,956 total reads. The genome harbored 5,796 coding sequences (CDSs), 76 tRNAs, and 5 rRNAs. A total of 4,835 CDSs were annotated with known functions, and 961 were hypothetical proteins. The plasmid was assembled in 8 contigs displaying 202,973 bp (N50 = 113,706 bp) and harbored 120 protein-coding genes and 154 hypothetical proteins. Phylogenetic analysis placed the isolate within the *K. variicola* clade with 100% bootstrap support (Fig. 1).

The bacterium harbors 27 putative antibiotic and biocide resistance genes, including multi-drug efflux pumps, as well as arsenic and multi-metal resistance genes. It contains 63 putative virulence genes distributed in adherence, iron uptake, secretion systems, and proteases, including a protease involved in the pathogenesis of *Yersinia pestis* (26). The strain also carries genes involved in nitrogen fixation, phosphate solubilization, and

phytohormone production typical in plant-interacting bacteria. The presence of genes associated with plant colonization and biocide resistance substantiates the possibility of pathogen transmission through plants to animals and humans.

ACKNOWLEDGMENTS

The authors would like to thank the funding agencies Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—CAPES, FAPESPA (Amazonia Foundation to Support Studies and Research), the CNPq (National Council for Scientific and Technological Development) project #312316/2022-4, Secretary of State for Science, Technology, and Professional and Technological Education (SECTET), Dean's Office for Research and Graduate Studies/Federal University of Pará—PROPESP/UFPA (PAPQ), and Ministerio de Educación Superior Ciencia y Tecnología de la República Dominicana (MESCYT) through the project: Mapeo de resistomas en los ríos Yaque del Norte y Yaque del Sur y sus implicaciones para la salud humana (FONDOCYT) for the financial support on this work.

This research was funded by the Ministerio de Educacion Superior Ciencia y Tecnologia (MESCYT) through Fondo Nacional de Innovación y Desarrollo Científico y Tecnológico (FONDOCYT), grant number 2020-2021-2C6-029:ABORDAJE DE ONEHEALTH PARA LA MEJORA DE LA CALIDAD E INNOCUIDAD DE LOS VEGETALES Y HORTALIZAS PRODUCIDOS Y COMERCIALIZADOS EN REPÚBLICA DOMINICANA A TRAVÉS DE LAS CIENCIAS OMICAS Y BIOINFORMÁTICA.

AUTHOR AFFILIATIONS

¹Laboratory of Bioinformatic and Genomics of Microorganisms, Institute of Biological Sciences, Federal University of Pará-UFPA, Belém, Pará, Brazil

²Centro Paraense de Computação Distribuída e de Alto Desempenho, Federal University of Para, Belém, Pará, Brazil

³Institute of Biological Sciences, Universidade Federal de Minas Gerais, Belo Horizonte, Minas Gerais, Brazil

⁴Laboratório de Genética, Evolução e Bioinformática, Instituto Federal do Pará-IFPA, Tucuruí, Pará, Brazil

⁵Laboratory of Applied Biology and Bioinformatics, Department of Basic and Environmental Sciences, Instituto Tecnológico de Santo Domingo (INTEC), Santo Domingo, Dominican Republic

⁶Universidad Tecnológica de Santiago, Campus Central de Herrera, Santo Domingo, Dominican Republic

⁷Instituto de Innovación en Biotecnología e Industria (IIBI), Santo Domingo, Distrito Nacional, Dominican Republic

AUTHOR ORCIDs

David Tavares Martins  <http://orcid.org/0009-0008-2595-8756>

Edian Franco Franklin de Los Santos  <http://orcid.org/0000-0001-9715-9437>

Rommel Thiago Jucá Ramos  <http://orcid.org/0000-0002-8032-1474>

AUTHOR CONTRIBUTIONS

David Tavares Martins, Conceptualization, Data curation, Investigation, Methodology, Writing – original draft, Writing – review and editing | Oscar Victor Cardenas Alegria, Investigation, Methodology | Thais Azevedo Cardoso, Methodology | Antônio Márcio Gomes Martins-Júnior, Investigation, Methodology | Luiz Orlando Maroto-Martin, Investigation, Methodology | Edian Franco Franklin de Los Santos, Funding acquisition, Investigation, Methodology, Project administration | Rommel Thiago Jucá Ramos, Project administration, Supervision, Writing – review and editing

DATA AVAILABILITY

This Whole Genome Shotgun project has been deposited in GenBank under the bioproject [PRJNA951268](https://www.ncbi.nlm.nih.gov/bioproject/PRJNA951268) and biosample [SAMN49305619](https://www.ncbi.nlm.nih.gov/biosample/SAMN49305619). The raw data are available at [SRR34021923](https://www.ncbi.nlm.nih.gov/srr/SRR34021923) and the assembly at [GCF_051107315.1](https://www.ncbi.nlm.nih.gov/genbank/GCF_051107315.1). Supplementary information can be found in [10.5281/zenodo.17019365](https://doi.org/10.5281/zenodo.17019365).

REFERENCES

- Yang L, Yang K. 2020. Biological function of *Klebsiella variicola* and its effect on the rhizosphere soil of maize seedlings. *PeerJ* 8:e9894. <https://doi.org/10.7717/peerj.9894>
- Rodríguez-Medina N, Barrios-Camacho H, Duran-Bedolla J, Garza-Ramos U. 2019. *Klebsiella variicola*: an emerging pathogen in humans. *Emerg Microbes Infect* 8:973–988. <https://doi.org/10.1080/22221751.2019.1634981>
- Afzal AM, Rasool MH, Waseem M, Aslam B. 2017. Assessment of heavy metal tolerance and biosorptive potential of *Klebsiella variicola* isolated from industrial effluents. *AMB Express* 7:184. <https://doi.org/10.1186/s13568-017-0482-2>
- Mahgoub SA, Qattan SYA, Salem SS, Abdelbasit HM, Raafat M, Ashkan MF, Al-Quwaie DA, Motwali EA, Alqahtani FS, Abd El-Fattah HI. 2023. Characterization and biodegradation of phenol by *Pseudomonas aeruginosa* and *Klebsiella variicola* strains isolated from sewage sludge and their effect on soybean seeds germination. *Molecules* 28:1203. <https://doi.org/10.3390/molecules28031203>
- Zhang J, Liang S, Wang X, Lu Z, Sun P, Zhang H, Sun F. 2019. Biodegradation of Atrazine by the Novel *Klebsiella variicola* Strain FH-1. *Biomed Res Int* 2019:1–12. <https://doi.org/10.1155/2019/4756579>
- de Campos TA, de Almeida FM, de Almeida APC, Nakamura-Silva R, Oliveira-Silva M, de Sousa IFA, Cerdeira L, Lincopan N, Pappas GJ Jr, Pitondo-Silva A. 2021. Multidrug-resistant (MDR) *Klebsiella variicola* strains isolated in a Brazilian Hospital belong to new clones. *Front Microbiol* 12:604031. <https://doi.org/10.3389/fmicb.2021.604031>
- Potter RF, Lainhart W, Twentyman J, Wallace MA, Wang B, Burnham CAD, Rosen DA, Dantas G. 2018. Population structure, antibiotic resistance, and uropathogenicity of *Klebsiella variicola*. *mBio* 9:e02481-18. <https://doi.org/10.1128/mBio.02481-18>
- Huang L, Fu L, Hu X, Liang X, Gong G, Xie C, Zhang F, Wang Y, Zhou Y. 2021. Co-occurrence of *Klebsiella variicola* and *Klebsiella pneumoniae* both carrying *bla_{KPC}* from a respiratory intensive care unit patient. *Infect Drug Resist* 14:4503–4510. <https://doi.org/10.2147/IDR.S330977>
- Lindgreen S. 2012. AdapterRemoval: easy cleaning of next-generation sequencing reads. *BMC Res Notes* 5:337. <https://doi.org/10.1186/1756-0500-5-337>
- Melsted P, Halldórsson BV. 2014. KmerStream: streaming algorithms for *k*-mer abundance estimation. *Bioinformatics* 30:3541–3547. <https://doi.org/10.1093/bioinformatics/btu713>
- Andrews S. 2023. FastQC: A quality control analysis tool for high throughput sequencing data. Available from: <https://github.com/s-andrews/FastQC>
- Schmidt B, Sinha R, Beresford-Smith B, Puglisi SJ. 2009. A fast hybrid short read fragment assembly algorithm. *Bioinformatics* 25:2279–2280. <https://doi.org/10.1093/bioinformatics/btp374>
- Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Pribelski 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>
- Wick RR, Judd LM, Gorrie CL, Holt KE. 2017. Unicycler: resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol* 13:e1005595. <https://doi.org/10.1371/journal.pcbi.1005595>
- 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>
- Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069. <https://doi.org/10.1093/bioinformatics/btu153>
- Robertson J, Nash JHE. 2018. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. *Microb Genom* 4:e000206. <https://doi.org/10.1099/mgen.0.000206>
- Miranda FM, Ramos RTJ. 2025. labgm/assembly-hiseq: Initial release (v1.0). Zenodo. Available from: <https://doi.org/10.5281/zenodo.16941612>
- Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2022. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics* 38:5315–5316. <https://doi.org/10.1093/bioinformatics/btad672>
- Gillespie JJ, Wattam AR, Cammer SA, Gabbard JL, Shukla MP, Dalay O, Driscoll T, Hix D, Mane SP, Mao C, Nordberg EK, Scott M, Schulman JR, Snyder EE, Sullivan DE, Wang C, Warren A, Williams KP, Xue T, Seung Yoo H, Zhang C, Zhang Y, Will R, Kenyon RW, Sobral BW. 2011. PATRIC: the comprehensive bacterial bioinformatics resource with a focus on human pathogenic species. *Infect Immun* 79:4286–4298. <https://doi.org/10.1128/IAI.00207-11>
- Bonin N, Doster E, Worley H, Pinnell LJ, Bravo JE, Ferm P, Marini S, Prosperi M, Noyes N, Morley PS, Boucher C. 2023. MEGARes and AMR++, v3.0: an updated comprehensive database of antimicrobial resistance determinants and an improved software pipeline for classification using high-throughput sequencing. *Nucleic Acids Res* 51:D744–D752. <https://doi.org/10.1093/nar/gkac1047>
- Alcock BP, Huynh W, Chalil R, Smith KW, Raphenya AR, Wlodarski MA, Edalatmand A, Petkau A, Syed SA, Tsang KK, et al. 2023. CARD 2023: expanded curation, support for machine learning, and resistome prediction at the Comprehensive Antibiotic Resistance Database. *Nucleic Acids Res* 51:D690–D699. <https://doi.org/10.1093/nar/gkac920>
- Liu B, Zheng D, Jin Q, Chen L, Yang J. 2019. VFDB 2019: a comparative pathogenomic platform with an interactive web interface. *Nucleic Acids Res* 47:D687–D692. <https://doi.org/10.1093/nar/gky1080>
- 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>
- Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30:1312–1313. <https://doi.org/10.1093/bioinformatics/btu033>
- Demeure CE, Dussurget O, Mas Fiol G, Le Guern AS, Savin C, Pizarro-Cerdá J. 2019. *Yersinia pestis* and plague: an updated view on evolution, virulence determinants, immune subversion, vaccination, and diagnostics. *Genes Immun* 20:357–370. <https://doi.org/10.1038/s41435-019-0065-0>